

Achieving the Holevo bound via a bisection decoding protocol

Matteo Rosati and Vittorio Giovannetti¹

*NEST, Scuola Normale Superiore and Istituto di Nanoscienze-CNR,
Piazza dei Cavalieri 7, I-56126 Pisa, Italy*

We present a new decoding protocol to realize transmission of classical information through a quantum channel at asymptotically maximum capacity, achieving the Holevo bound and thus the optimal communication rate. At variance with previous proposals, our scheme recovers the message bit by bit, making use of a series “yes-no” measurements, organized in bisection fashion, thus determining which codeword was sent in $\log_2 N$ steps, N being the number of codewords.

I. INTRODUCTION

One of the main achievements in quantum information theory has been the development of a generalization of Shannon's theory for quantum communication¹. In particular, the Holevo bound^{2,3} sets a limit on the rate of reliable transmission of classical information through a quantum channel, which is also achievable in the asymptotic limit of infinitely long sequences^{4–15}. Consequently, via proper optimization and regularization¹⁶, it provides the quantum analog of the Shannon classical capacity formula.

The original proof^{4,5} was carried out by extending to the quantum regime the concept of typical subspaces used in Shannon communication theory^{17,18}. A crucial point is the choice of a proper POVM which allows Bob to identify the right message with small error probability. The first explicit detection scheme used in this context is a one-step collective-measurement POVM known as *Pretty Good Measurement* (PGM)^{5,6}, highly effective theoretically but not easily realizable in practice.

Following the proof of Ogawa and Nagaoka^{9,10}, Hayashi and Nagaoka¹¹, which establishes a connection with the quantum-hypothesis-testing problem¹⁹, the possibility of asymptotically achieving the bound through a series of “yes-no” projective measurements was investigated^{13–15}. This sequential protocol checks whether the received state resides in the typical subspace of a given codeword, for each codeword in the code, until it receives a positive answer or else declares failure. The “yes-no” question is asked, for each codeword, by applying the projector on its typical subspace and thus makes the decoding protocol more suited for practical implementations than the PGM. Indeed a design for an explicit and structured optical receiver was proposed^{20,21}, which used this protocol, with applications both to optical communication and quantum reading. In particular, for a lossy bosonic channel²² (a model most commonly used to represent realistic fiber and free-space communication) it was shown that the sequential decoder can be built with gaussian displacement operators and vacuum-or-not measurements^{15,19,23}. An alternative, near-explicit approach, for capacity-achieving classical-quantum communication was also recently developed by Wilde and Guha²⁴, adapting to the quantum scenario the classical polar coding introduced by Arikan²⁵. In particular, making use of optimal Helstrom measurements in the quantum-hypothesis-testing procedure and of Sen's non-commutative union bound¹⁵, they proposed an encoding technique which realizes channel polarization and consequently introduced a

quantum successive cancellation decoder. Later work modified such decoding strategy to obtain a partially non-collective measurement²⁶ and extended polar coding to private and quantum communication through arbitrary qubit channels^{27–29}. The relevance of this approach is associated with the fact that, at variance with other proposals^{4,5,13–15}, it allows optimal decoding with a linear (in the amount of bits) number of collective measurements.

In this paper we propose a bisection decoding scheme for classical communication through a quantum channel and show that it achieves the maximum capacity in the asymptotic limit of infinitely long codewords, providing yet an alternative proof of the attainability of the Holevo bound. While being inspired to the sequential decoding algorithm^{13–15}, analogously to Refs. 20 and 21 our scheme exhibits an exponential advantage in the number of measurements which have to be performed in order to recover the message: specifically if the sequential method is built on $O(N)$ concatenated “yes-no” detections, where N is the number of codewords, the bisection method only requires $\log_2 N$ of such “atomic” steps, thus scaling linearly with the number of bits n which one wishes to transmit. We stress however that, being our individual detections explicitly many-body operations, at present we have no evidence in support of the fact that such advantage could be translated in a decoding scheme which is efficient from the computational point of view, i.e. in terms of the number of quantum gates one has to apply to the received string of quantum information carriers. A similar problem arises also in the case of polar codes (see e.g. Ref. 26), and it is caused by the lack of an explicit implementation (or at least of an estimate of its complexity) of the “atomic” steps involved in the two protocols, i.e. the “yes-no” set detections for the present method and the Helstrom measurement for polar coding. Still we believe that our method can be of some interest as it widens the class of known decoding strategies which are asymptotically optimal, increasing hence the chances of identifying at least one which is suitable to implementations. In this respect it is also worth noticing that the proposed scheme exhibits the nontrivial advantage of gaining a bit of information at each step of the procedure, a feature which may be extremely appealing when dealing with faulty decoders, as it allows partial identification of the transmitted message even in the presence of subsequent detection failures.

As in all the previous works on the subject, in our derivation we heavily rely on the structure of typical projectors, although we need to properly combine them in order to build

efficient “yes-no” set measurements which reconstruct the message bit-by-bit by checking, at each step of the procedure, whether the received message belongs to one of two possible sets of codewords. In an effort to make the paper self-contained, we reproduce a series of known results^{1,30} providing, in some cases, alternative proofs which are explicitly presented in the framework which best fits with the proposed approach.

The paper is organized as follows: we start in Sec. II, where we introduce the notation and state the problem in a rigorous way. In Sec. III we present some mathematical tools which are important to derive our results. In particular Sec. III A is devoted to review some basic facts about the structure of typical subspaces of a quantum source, while Sec. III B discusses few Lemmas which allow us to put bounds on the probability of retrieving certain POVM outcomes from states which are close to each other. The bisection protocol is introduced in Sec. IV, identifying a sufficient condition which ensures it can asymptotically attain the Holevo bound in Sec. IV B and presenting three different methods which satisfy this condition. Conclusions are finally given in Sec. V.

II. THE PROBLEM: ACHIEVING THE HOLEVO BOUND

Consider a memoryless quantum communication channel described by a completely positive, trace preserving (CPT) mapping² \mathcal{T} that Alice (the sender of the communication scheme) uses to transmit classical messages to Bob (the receiver). Given an alphabet \mathcal{A} of classical symbols, we define a N -element code $\mathcal{C} := \{\vec{j}^{(1)}, \dots, \vec{j}^{(N)}\}$ as a subset of \mathcal{A}^n which contains N selected n -long strings $\vec{j} := (j_1, \dots, j_n)$ of elements of \mathcal{A} : they represent the codewords which are employed by Alice to codify N distinct classical messages. A quantum encoding is then realized by assigning a mapping which, given $j \in \mathcal{A}$, associates to it a density matrix $\sigma_j \in \mathfrak{S}(\mathcal{H})$ of the quantum carrier that propagates through the channel. Accordingly each string $\vec{j} \in \mathcal{A}^n$ will be represented by the product state $\sigma_{\vec{j}} := \sigma_{j_1} \otimes \dots \otimes \sigma_{j_n} \in \mathfrak{S}(\mathcal{H}^{\otimes n})$, and received by Bob as

$$\rho_{\vec{j}} := \rho_{j_1} \otimes \dots \otimes \rho_{j_n}, \quad (1)$$

where $\rho_j := \mathcal{T}[\sigma_j]$ is the output density matrix corresponding to the input σ_j . In this framework each classical code \mathcal{C} is associated with a quantum code via the following classical-

to-quantum correspondence

$$\mathcal{C} = \{\vec{j}^{(1)}, \dots, \vec{j}^{(N)}\} \longrightarrow \mathbf{C} := \{\rho_{\vec{j}^{(1)}}, \dots, \rho_{\vec{j}^{(N)}}\}, \quad (2)$$

the states $\rho_{\vec{j}^{(\ell)}}$ being those which Bob has to discriminate in order to recover the message Alice sent to him while using the code \mathcal{C} . For such purpose he will employ a decoding POVM of elements

$$\left\{ X_1, \dots, X_N, X_0 = \mathbf{1} - \sum_{\ell=1}^N X_\ell \right\}, \quad (3)$$

whose outcome represents the inferred value of the transmitted message. Specifically for $\ell = 1, \dots, N$, the operator X_ℓ is associated with the event where Bob assumes that the received message is the ℓ -th one, while X_0 is associated with an explicit failure of the decoding stage. Accordingly the average error probability of the quantum code \mathbf{C} can then be computed as

$$P_{err}(\mathbf{C}) := \frac{1}{N} \sum_{\ell=1}^N [1 - p_{succ}(\ell)] = 1 - \frac{1}{N} \sum_{\ell=1}^N p_{succ}(\ell). \quad (4)$$

where

$$p_{succ}(\ell) := \text{Tr} \left[X_\ell \rho_{\vec{j}^{(\ell)}} \right], \quad (5)$$

is the probability that Bob will successfully retrieve the ℓ -th codeword when Alice transmits it.

In the long message limit $n \rightarrow \infty$, it has been shown³⁻⁵ that $P_{err}(\mathbf{C})$ can be sent to zero if the number of messages scales as $N = 2^{nR}$, R being the transmission rate of the scheme which is bounded by the Holevo theorem. Specifically we must have that

$$R \leq \max_{\{p_j, \sigma_j\}} \chi(\{p_j, \rho_j\}) = C_{Hol}, \quad (6)$$

where on the right-hand-side the maximization is performed over all possible input ensembles $\{p_j, \sigma_j : j \in \mathcal{A}\}$ obtained by selecting the state σ_j with probability distribution p_j , and where the Holevo information of the associated output ensemble $\{p_j, \rho_j = \mathcal{T}(\sigma_j); j \in \mathcal{A}\}$ is defined as

$$\chi(\{p_j, \rho_j\}) := S \left(\sum_j p_j \rho_j \right) - \sum_j p_j S(\rho_j), \quad (7)$$

by means of the Von Neumann entropy $S(\rho) = -\text{Tr}[\rho \log_2 \rho]$.

It is known that the inequality (6) is achievable, in the sense that, for any output ensemble $\mathcal{E} := \{p_j, \rho_j; j \in \mathcal{A}\}$, one can identify a set \mathbf{C} of $N \sim 2^{n \chi(\{p_j, \rho_j\})}$ quantum codewords and a

decoding POVM (3) for which the error probability (4) goes to zero as n increases. This can be done by exploiting what, in classical information theory, is known as Shannon's averaging trick. The idea is as follows: the ensemble \mathcal{E} can be seen as a source which, when operating n times, will produce n -long product states $\rho_{\vec{j}}$ of the form (1) with probability

$$p_{\vec{j}} = p_{j_1} p_{j_2} \cdots p_{j_n}. \quad (8)$$

Therefore iterating N times this operation, \mathcal{E} will be able to generate a code \mathbf{C} defined as in Eq. (2) with probability

$$P(\mathbf{C}) = \prod_{\ell=1, \dots, N} p_{\vec{j}^{(\ell)}} = \prod_{\ell=1, \dots, N} \prod_{q=1}^n p_{j_q^{(\ell)}}, \quad (9)$$

where $\vec{j}^{(\ell)}$ are the codewords of the classical counterpart \mathcal{C} of \mathbf{C} . The set $\mathcal{S} := \{\mathbf{C}, P(\mathbf{C})\}$ defines the statistical collection of the quantum codes one can associate to \mathcal{E} for fixed values of N and n . Accordingly, instead of optimizing the total error probability (4) of a single element of such a set, we can now consider its averaged value with respect to the probability $P(\mathbf{C})$,

$$\langle P_{err} \rangle_{\mathcal{S}} := \sum_{\mathbf{C}} P(\mathbf{C}) P_{err}(\mathbf{C}) = 1 - \frac{1}{N} \sum_{\ell=1}^N \langle p_{succ}(\ell) \rangle_{\mathcal{S}}, \quad (10)$$

the rationale being that if this quantity can be forced to go to zero in the limit $n \rightarrow \infty$ then at least one (actually almost all) code exists in \mathcal{S} for which $P_{err}(\mathbf{C})$ tends to zero in the same limit.

The first proof^{4,5} of this fact made use of a single-step decoding POVM (3), known as *pretty good measurement* (PGM) or *square root measurement*, which is extremely efficient from a theoretical point of view but difficult to implement. More recently, a sequential decoding scheme has been introduced^{13–15}, which makes use of projective “yes-no” measurements to verify whether the received state corresponds to a certain codeword or not. Following an arbitrary ordering of codewords, this question is asked for each of them in turn, until either a positive answer is obtained for some \vec{j} or else a negative answer for all the codewords. To some extent the sequential scheme appears to be easier to realize in practice as it decomposes the process into a series of simple steps, and indeed several proposals have been made for its use in the context of continuous variable communication lines^{22,32–34}. Still it has a major drawback in its scaling, since an order of $N = 2^{nR}$ operations is required for its application. The protocol presented here is inspired by the sequential decoding but

makes use of a bisection method, performing at each step a “yes-no” measurement for a set of possible codewords, whose size is progressively halved, allowing Bob to recover the transmitted message bit-by-bit.

III. MATHEMATICAL TOOLS

This section reviews some basic facts about typical subspaces and presents some inequalities which will be useful in proving the optimality of our decoding scheme. For a complete description of the following properties we refer the reader to Refs. 1, 2, 15, 35, and 36.

A. Typical subspaces

Consider the average state

$$\rho = \sum_{j \in \mathcal{A}} p_j \rho_j = \sum_x q_x |e_x\rangle\langle e_x|, \quad (11)$$

of the quantum source $\mathcal{E} := \{p_j, \rho_j; j \in \mathcal{A}\}$ and its spectral decomposition in terms of the eigenbasis $\{|e_x\rangle\}$ of \mathcal{H} and the eigenvalues $\{q_x\}$. This induces a classical random variable X with probability distribution q_x which, on n sampling events, produces the sequence $\vec{x} = (x_1, \dots, x_n)$ with probability $q_{\vec{x}} = \prod_{\ell=1}^n q_{x_\ell}$. The classical δ -typical subspace T_δ^n is defined as the subspace of such sequences whose sample entropy differs from the expected entropy of the random variable for less than a given quantity $\delta > 0$:

$$T_\delta^n = \{\vec{x} : |\bar{H}(\vec{x}) - H(X)| \leq \delta\}, \quad (12)$$

the sample entropy of a codeword being

$$\bar{H}(\vec{x}) = -\frac{1}{n} \log_2 q_{\vec{x}} = -\frac{1}{n} \sum_{i=1}^n \log_2 q_{x_i}, \quad (13)$$

i.e. the average information content of the n symbols in the \vec{x} sequence, while the associated Shannon entropy is defined as usual:

$$H(X) = -\sum_x q_x \log_2 q_x = S(\rho), \quad (14)$$

where in the last identity we used the correspondence with the von Neumann entropy functional of the average state ρ . As a consequence, the δ -typical subspace $\mathcal{H}_{typ}^{(n)}$ of quantum

state ρ is made of all those vectors $|e_{\vec{x}}\rangle$ whose corresponding classical sequence is δ -typical, i.e. $\vec{x} \in T_{\delta}^n$. The projector on this subspace is given by

$$P = \sum_{\vec{x} \in T_{\delta}^n} |e_{\vec{x}}\rangle \langle e_{\vec{x}}|. \quad (15)$$

Similar properties as for the classical typical subspace hold for the quantum one, namely

$$\text{Tr} [P \rho^{\otimes n}] \geq 1 - \epsilon_1, \quad (16)$$

$$\text{Tr} [P] \leq 2^{n[S(\rho)+\delta]}, \quad (17)$$

$$2^{-n[S(\rho)+\delta]} P \leq P \rho^{\otimes n} P \leq 2^{-n[S(\rho)-\delta]} P, \quad (18)$$

for $\epsilon_1 > 0$ and n sufficiently large. These properties state respectively that:

- The quantum state $\rho^{\otimes n}$ resides with high probability in the δ -typical subspace of ρ ;
- The size of the δ -typical subspace is exponentially smaller than the size of the whole space, unless the source is maximally mixed, i.e. $S(\rho) = \log_2 d$;
- The probability distribution of δ -typical sequences is approximately uniform $\sim 2^{-nS(\rho)}$.

It is finally important to observe that the parameter ϵ_1 entering in Eq. (16) can be linked to n via an exponential scaling³⁷, i.e. $\epsilon_1 = O(e^{-n})$, which ensures that for all polynomial functions $\text{poly}(n)$ of n one has

$$\lim_{n \rightarrow \infty} \text{poly}(n) \epsilon_1 = 0, \quad (19)$$

(see Appendix A for details).

Similar typical subspaces can be identified also for each specific state $\rho_{\vec{j}}$ produced by the source, i.e. for each codeword in \mathbf{C} , by using the notion of conditional typicality. Indeed each source state can be seen as a classical-quantum state $|j\rangle\langle j| \otimes \rho_j$ and its spectral decomposition will be in terms of eigenvectors $\{|j\rangle \otimes |e_y^j\rangle\}$ and eigenvalues $\{\lambda_y^j\}$. This again induces the classical random variables J , with probability distribution p_j representing the possible states emitted by the source, and Y , with conditional probability distribution $\lambda_y^j = p(y|j)$. The classical δ -conditionally typical subspace is then defined for each n -long sequence \vec{j} as

$$T_{\delta}^{\vec{j}} = \left\{ \vec{y} : |\bar{H}(\vec{y}|\vec{j}) - H(Y|J)| \leq \delta \right\}, \quad (20)$$

where now the entropic quantities are conditional ones, i.e.

$$\bar{H}(\vec{y}|\vec{j}) = -\frac{1}{n} \log_2 \lambda_{\vec{y}}^{\vec{j}} = -\frac{1}{n} \sum_{i=1}^n \log_2 \lambda_{y_i}^{j_i} \quad (21)$$

$$H(Y|J) = \sum_{j \in \mathcal{A}} p_j H(Y|j) = -\sum_{j,y} p_j \lambda_y^j \log_2 \lambda_y^j. \quad (22)$$

The δ -conditionally typical subspace $\mathcal{H}_{typ}^{\vec{j}}$ of quantum codeword state $\rho_{\vec{j}}$ is made of all those vectors $|e_{\vec{y}}^{\vec{j}}\rangle$ whose corresponding classical sequence is δ -conditionally typical, i.e. $\vec{y} \in T_{\delta}^{\vec{j}}$.

The projector on this subspace is given by

$$P_{\vec{j}} = \sum_{\vec{y} \in T_{\delta}^{\vec{j}}} |e_{\vec{y}}^{\vec{j}}\rangle \langle e_{\vec{y}}^{\vec{j}}|. \quad (23)$$

Given $\epsilon_2 > 0$ and n sufficiently large, the following three main properties hold for the conditionally typical subspace:

$$\sum_{\vec{j}} p_{\vec{j}} \text{Tr} [P_{\vec{j}} \rho_{\vec{j}}] \geq 1 - \epsilon_2, \quad (24)$$

$$\sum_{\vec{j}} p_{\vec{j}} \text{Tr} [P_{\vec{j}}] \leq 2^n [\sum_{j \in \mathcal{A}} p_j S(\rho_j) + \delta], \quad (25)$$

$$2^{-n [\sum_{j \in \mathcal{A}} p_j S(\rho_j) + \delta]} P_{\vec{j}} \leq P_{\vec{j}} \rho_{\vec{j}} P_{\vec{j}} \leq 2^{-n [\sum_{j \in \mathcal{A}} p_j S(\rho_j) - \delta]} P_{\vec{j}}, \quad (26)$$

where in the first two expressions the average³¹ is taken with respect to the joint probability $p_{\vec{j}}$ of \mathcal{E} introduced in Eq. (8), while the last inequality applies for all \vec{j} . As for Eq. (16) we stress that the parameter ϵ_2 of Eq. (24) can be chosen to have an exponential scaling in n which guarantees that the condition (19) holds also in this case. Note finally that the conditionally typical subspaces of different codewords are in general not orthogonal, since they are built using vectors of two spectral decompositions of the same space $\mathcal{H}^{\otimes n}$.

B. Measurement Lemmas

We state here some Lemmas which will be used in the rest of the article. They relate in various ways quantum states before and after a measurement, with the slight but crucial detail that the latter need not be normalized. Formally one can represent them as *subnormalized* density matrices, i.e. positive operators whose trace is smaller than or equal to one.

An explicit proof of the first three Lemmas can be found in Appendix B: they refer to properties of the trace norm, which for a generic operator θ , is defined as $\|\theta\|_1 = \text{Tr}|\theta|$ with $|\theta| = \sqrt{\theta^\dagger \theta}$ being the modulus of θ . The last Lemma instead was proved by Sen¹⁵ and provides an alternative, useful, way of estimating the error probability of the sequential decoding protocol of Refs. 13 and 14.

Lemma 1. (*Measurement on approximately close states*) Let ρ, σ be subnormalized density matrices. Let E be a positive and less-than-one operator, i.e. $0 \leq E \leq \mathbf{1}$. Then

$$\text{Tr}[E\rho] \geq \text{Tr}[E\sigma] - 2D(\rho, \sigma), \quad (27)$$

where $D(\rho, \sigma) = \frac{1}{2} \|\rho - \sigma\|_1$ is the trace distance between ρ and σ .

Lemma 2. (*Gentle operator*) Let ρ be a subnormalized density matrix and E a positive and less-than-one operator, i.e. $0 \leq E \leq \mathbf{1}$. Let also $\langle \dots \rangle$ denote the average with respect to some probability distribution, which ρ and E may depend on. Suppose that, for some $1 \geq \epsilon > 0$,

$$\langle \text{Tr}[E\rho] \rangle \geq 1 - \epsilon. \quad (28)$$

Then

$$\left\langle D\left(\sqrt{E}\rho\sqrt{E}, \rho\right) \right\rangle \leq \sqrt{\epsilon}. \quad (29)$$

The two previous lemmas are well known for ordinary density matrices; they can be proved also for subnormalized ones by use of the following lemma.

Lemma 3. (*Alternative form of trace norm for subnormalized states*) Let ω be a hermitian operator (in particular, ω could be a subnormalized density matrix). Then

$$\|\omega\|_1 = \max_{-\mathbf{1} \leq \Lambda \leq \mathbf{1}} \text{Tr}[\Lambda\omega]. \quad (30)$$

Lemma 4. (*Contractivity of trace distance for POVM elements*) Let ρ, σ be subnormalized density matrices and $0 \leq E \leq \mathbf{1}$ a positive and less-than-one operator (for example it could be a POVM element and/or a projector). Then

$$D(E\rho E, E\sigma E) \leq D(\rho, \sigma). \quad (31)$$

Proof. Consider the expression of the trace norm of a hermitian operator as in Lemma 3 and apply it to the LHS of (31):

$$2D(E\rho E, E\sigma E) = \max_{-1 \leq \Lambda \leq 1} \text{Tr} [\Lambda E(\rho - \sigma)E] \quad (32)$$

$$= \text{Tr} [\bar{\Lambda} E(\rho - \sigma)E] = \text{Tr} [\Lambda'(\rho - \sigma)] \quad (33)$$

$$\leq \max_{-1 \leq \Lambda \leq 1} \text{Tr} [\Lambda(\rho - \sigma)] = 2D(\rho, \sigma). \quad (34)$$

The second equality follows from explicitly using the operator $\bar{\Lambda}$ which attains the maximum in (32). The third equality follows from using the cyclic property of the trace and setting $\Lambda' = E\bar{\Lambda}E$. The inequality follows from the fact that also Λ' is positive and less-than-one. \square

Lemma 5. (*Sen's Lemma*) *Let ρ be a subnormalized density matrix and P_1, \dots, P_k orthogonal projectors on subspaces of its Hilbert space. Let also $Q_i = \mathbf{1} - P_i$ be their complementary projectors. Then*

$$\text{Tr} [P_k \dots P_1 \rho P_1 \dots P_k] \geq \text{Tr} [\rho] - 2 \sqrt{\sum_{i=1}^k \text{Tr} [\rho Q_i]}. \quad (35)$$

IV. THE BISECTION PROTOCOL

A. Description of the protocol

In this subsection we introduce our decoding protocol (Fig. 1) which, given a density matrix extracted from a $N = e^{nR}$ -element quantum code \mathbf{C} (2), generated by the source \mathcal{E} , tries to identify it by using a bisection method. The measurement process comprises of $u_F = nR$ nested detection events, each aimed to recover one bit of information from the transmitted signal.

As a preliminary step, Bob assigns an ordering of the codewords in \mathbf{C} , identifying each of them with a unique string of u_F bits, $\vec{k} = (k_1, k_2, \dots, k_{u_F})$, e.g. by providing a binary representation of their label $\ell \in \{1, \dots, N\}$. In particular the first bit of the string \vec{k} identifies two distinct subsets of \mathbf{C} containing each $N/2$ codewords: the subset $\mathbf{C}_0^{(1)}$ formed by the codewords whose corresponding strings start with $k_1 = 0$, and the subset $\mathbf{C}_1^{(1)}$ characterized by those for which instead $k_1 = 1$. The second bit of the string \vec{k} is then used to further halve $\mathbf{C}_0^{(1)}$ and $\mathbf{C}_1^{(1)}$. Specifically for $k_1 = 0, 1$, $\mathbf{C}_{k_1}^{(1)}$ is split into the sub-subsets $\mathbf{C}_{k_1, k_2=0}^{(2)}$

and $\mathbf{C}_{k_1, k_2=1}^{(2)}$ which includes the $N/4$ codewords whose bits strings have k_1 as first bit and $k_2 = 0$ and $k_2 = 1$ as second bit, respectively. Proceeding along the same line Bob identifies hence a hierarchy of subsets organized in u_F sets, the u -th one being composed by 2^u disjoint subsets $\mathbf{C}_{k_1, k_2, \dots, k_u}^{(u)}$ labelled by the indexes k_1, k_2, \dots, k_u , and containing each $2^{u_F-u} = N/2^u$ codewords. Specifically $\mathbf{C}_{k_1, k_2, \dots, k_u}^{(u)}$ is the set formed by the codewords whose identifier string \vec{k} admits the value k_1 as first bit, the value k_2 as second bit, \dots , and the value k_u as the u -th bit. By construction for all $u \in \{1, \dots, u_F\}$ they fulfill the identities

$$\mathbf{C}_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)} \cap \mathbf{C}_{k_1, k_2, \dots, k_{u-1}, 1}^{(u)} = \emptyset, \quad (36)$$

$$\mathbf{C}_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)} \cup \mathbf{C}_{k_1, k_2, \dots, k_{u-1}, 1}^{(u)} = \mathbf{C}_{k_1, k_2, \dots, k_{u-1}}^{(u-1)}, \quad (37)$$

and the completeness relation

$$\mathbf{C} = \bigcup_{k_1, k_2, \dots, k_u \in \{0, 1\}} \mathbf{C}_{k_1, k_2, \dots, k_u}^{(u)}. \quad (38)$$

To recover which codeword Alice is transmitting, Bob performs a sequence of u_F concatenated measurements organized as shown in Fig. 1. The first of these measures is aimed to determine the value of the first bit k_1 of the bit string associated with the transmitted codeword, i.e. it allows Bob to determine whether the codeword is in the subset $\mathbf{C}_0^{(1)}$ or in the subset $\mathbf{C}_1^{(1)}$. The exact form of such procedure will be assigned in the following sections where three alternative examples of the scheme will be discussed in details: for the moment it is sufficient to observe that it can be described as a POVM $\mathcal{M}^{(1)}$ of elements $N_0^{(1)}, N_1^{(1)}$ associated respectively to the outcomes $k_1 = 0$ and $k_1 = 1$, plus a null term $N_{null}^{(1)} = \mathbf{1} - N_0^{(1)} - N_1^{(1)}$ associated with the case in which no decision can be made on the value of k_1 : if this event occurs simply Bob declares failure of the decoding procedure and stops the protocol (in the first implementation of the scheme we discuss in Sec. IV B 1 this element is not present, which is equivalent to set $N_{null}^{(1)} = 0$). Once k_1 has been determined, Bob proceeds with the second step of the protocol aimed to recover the value of the bit k_2 of the transmitted codeword. To this purpose, conditioned on the value of $k_1 \in \{0, 1\}$ obtained in the previous step, Bob performs now a new POVM $\mathcal{M}_{k_1}^{(2)}$ aimed to determine whether the received codeword belongs to $\mathbf{C}_{k_1, 0}^{(2)}$ or to $\mathbf{C}_{k_1, 1}^{(2)}$. Also $\mathcal{M}_{k_1}^{(2)}$ is characterized by three elements: $N_{k_1, 0}^{(2)}, N_{k_1, 1}^{(2)}$ corresponding to the cases $k_2 = 0$ and $k_2 = 1$ respectively, and $N_{k_1, null}^{(2)} = \mathbf{1} - N_{k_1, 0}^{(2)} - N_{k_1, 1}^{(2)}$ corresponding to the failure event (again the explicit expressions for these operators will be assigned later on). The procedure iterates till Bob either gets a failure event or recovers all

the u_F bits which identify the transmitted codeword. Specifically, assuming that no failures have occurred in the first $u - 1$ steps yielding the values k_1, k_2, \dots, k_{u-1} for the associated bits, at the u -th step Bob performs on the system a POVM $\mathcal{M}_{k_1, k_2, \dots, k_{u-1}}^{(u)}$ of elements $N_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)}$, $N_{k_1, k_2, \dots, k_{u-1}, 1}^{(u)}$, and $N_{k_1, k_2, \dots, k_{u-1}, \text{null}}^{(u)} = \mathbf{1} - N_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)} - N_{k_1, k_2, \dots, k_{u-1}, 1}^{(u)}$ to decide whether the received codeword belongs to set $\mathbf{C}_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)}$ or to $\mathbf{C}_{k_1, k_2, \dots, k_{u-1}, 1}^{(u)}$.

Given the above construction the probability of recovering a given string of bits $\vec{k} = (k_1, k_2, \dots, k_{u_F})$ when measuring an input state $\rho_{\vec{j}} \in \mathbf{C}$ can now be expressed along the lines detailed in Appendix C, i.e.

$$P(\vec{k}|\rho_{\vec{j}}) = \text{Tr}[F_{\vec{k}} \rho_{\vec{j}}], \quad (39)$$

with $F_{\vec{k}}$ defined as

$$F_{\vec{k}} = \left| \sqrt{N_{k_1, k_2, \dots, k_{u_F}}^{(u_F)}} \sqrt{N_{k_1, k_2, \dots, k_{u_F-1}}^{(u_F-1)}} \cdots \sqrt{N_{k_1, k_2}^{(2)}} \sqrt{N_{k_1}^{(1)}} \right|^2, \quad (40)$$

with $\{N_{k_1, \dots, k_u}^{(u)}\}_{u \in \{1, \dots, u_F = nR\}; k_1, \dots, k_{u-1} \in \{0, 1\}}$ the operators which define the POVM's of the protocol. The success probability of the procedure follows then from this expression by simply setting \vec{k} to coincide with the binary string associated with the selected \vec{j} , e.g. in the case of the ℓ -th codeword

$$p_{\text{succ}}(\ell) = P(\vec{k}^{(\ell)}|\rho_{\vec{j}^{(\ell)}}) = \text{Tr}[F_{\vec{k}^{(\ell)}} \rho_{\vec{j}^{(\ell)}}], \quad (41)$$

where $\vec{k}^{(\ell)}$ is the binary string corresponding to the index ℓ which defines the selected vector $\vec{j}^{(\ell)}$.

B. Attainability of the Holevo Bound via a bisection protocol

As detailed in the previous paragraphs a bisection protocol aimed to decode a $N = 2^{nR}$ -codewords quantum code \mathbf{C} is defined by assigning a family of three-outcomes POVM's $\mathcal{M}_{k_1, \dots, k_{u-1}}^{(u)}$ identified by the integer index $u \in \{1, \dots, u_F = nR\}$ and by the binary labels $k_1, \dots, k_{u-1} \in \{0, 1\}$ and concatenated as schematically shown in Fig. 1. In this section we are going to show that, as long as the rate R respects the inequality (6), it is possible to assign the operators $\{N_{k_1, \dots, k_u}^{(u)}\}_{u \in \{1, \dots, u_F = nR\}; k_1, \dots, k_{u-1} \in \{0, 1\}}$ which define the measures $\{\mathcal{M}_{k_1, \dots, k_{u-1}}^{(u)}\}_{u \in \{1, \dots, u_F = nR\}; k_1, \dots, k_{u-1} \in \{0, 1\}}$ in such a way that, in the limit of large n , the corresponding success probability (41) converges asymptotically to 1 when averaged over all

possible codes generated by the output source \mathcal{E} , i.e.

$$\lim_{n \rightarrow \infty} \langle p_{\text{succ}}(\ell) \rangle_{\mathcal{S}} = 1. \quad (42)$$

Accordingly the corresponding average error probability (10) asymptotically nullifies, proving hence that bisection decoding procedures can be used to saturate the Holevo bound. In order to achieve this goal we start by presenting a sufficient condition on $N_{k_1, \dots, k_u}^{(u)}$ which, if fulfilled, would yield the limit (42) independently of the value of R , see Theorem 1. Subsequently we show that for all rates R respecting the Holevo Bound (6) we can indeed fulfil such sufficient condition. This is done by presenting three independent choices of the operators $\{N_{k_1, \dots, k_u}^{(u)}\}_{u \in \{1, \dots, u_F = nR\}; k_1, \dots, k_{u-1} \in \{0, 1\}}$, corresponding to three different ways of constructing the bisection scheme: via orthogonal projections (see Sec. IV B 1); via PGM detections (see Sec. IV B 2); and via sequential detections (see Sec. IV B 3).

Theorem 1 (Sufficient Condition). *For n integer let \mathbf{C} be a quantum code formed by $N = 2^{nR}$ separable codewords (1) of length n extracted from the output ensemble \mathcal{E} , and a bisection protocol with POVM's $\{\mathcal{M}_{k_1, \dots, k_{u-1}}^{(u)}\}_{u \in \{1, \dots, u_F = nR\}; k_1, \dots, k_{u-1} \in \{0, 1\}}$ characterized by the operators $\{N_{k_1, \dots, k_u}^{(u)}\}_{u \in \{1, \dots, u_F = nR\}; k_1, \dots, k_{u-1} \in \{0, 1\}}$. The corresponding success probability (41) converges to one as in Eq. (42) when averaged over all possible codes generated by the output ensemble \mathcal{E} if, for all $\ell \in \{1, \dots, N\}$ and $u \in \{1, \dots, u_F\}$, one of the following conditions is fulfilled*

i)

$$\left\langle \text{Tr} \left[N_{k_1^{(\ell)}, \dots, k_u^{(\ell)}}^{(u)} \rho_{\vec{j}^{(\ell)}} \right] \right\rangle_{\mathcal{S}} \geq 1 - \epsilon(n); \quad (43)$$

ii)

$$\left\langle \text{Tr} \left[N_{k_1^{(\ell)}, \dots, k_u^{(\ell)}}^{(u)} P \rho_{\vec{j}^{(\ell)}} P \right] \right\rangle_{\mathcal{S}} \geq 1 - \epsilon(n), \quad (44)$$

with $\epsilon(n) > 0$ being a function that decreases asymptotically to zero faster than $1/n^2$ as n goes to infinity and P being the projector on the typical subspace of the average codeword associated with the output ensemble \mathcal{E} (in the above expression $k_1^{(\ell)}, \dots, k_u^{(\ell)}$ are the first u elements of the binary string $\vec{k}^{(\ell)}$ which represents the codeword index ℓ that labels the density matrix $\rho_{\vec{j}^{(\ell)}}$).

Proof. We start by directly proving that Eq. (43) is a sufficient condition for Eq. (42), part i) of the theorem. Part ii) of the theorem is then obtained by showing that Eq. (44) implies Eq. (43).

Part i): The success probability (5) that an element $\rho_{\vec{j}^{(\ell)}} \in \mathbf{C}$ will be correctly decoded by the bisection procedure characterized by the operators $\{N_{k_1, \dots, k_u}^{(u)}\}_{u \in \{1, \dots, u_F=nR\}; k_1, \dots, k_{u-1} \in \{0,1\}}$ can be computed as in Eq. (41) with $\vec{k}^{(\ell)}$ being the identifying bit string that Bob has assigned to the ℓ -th codeword. To put a bound on the average value of this quantity over the collection \mathcal{S} of quantum codes emitted by the source \mathcal{E} , we observe that

$$\begin{aligned} \langle p_{succ}(\ell) \rangle_{\mathcal{S}} &= \left\langle \text{Tr} \left[M_{u_F}^2 M_{u_F-1} \cdots M_1 \rho_{\vec{j}^{(\ell)}} M_1 \cdots M_{u_F-1} \right] \right\rangle_{\mathcal{S}} \\ &\geq \left\langle \text{Tr} \left[M_{u_F}^2 \rho_{\vec{j}^{(\ell)}} \right] - 2D \left(M_{u_F-1} \cdots M_1 \rho_{\vec{j}^{(\ell)}} M_1 \cdots M_{u_F-1}, \rho_{\vec{j}^{(\ell)}} \right) \right\rangle_{\mathcal{S}}, \end{aligned} \quad (45)$$

where for easiness of notation we introduced $M_u = \sqrt{N_{k_1^{(\ell)}, \dots, k_u^{(\ell)}}^{(u)}}$ and applied Lemma 1 with $E = M_{u_F}^2$, $\rho = M_{u_F-1} \cdots M_1 \rho_{\vec{j}^{(\ell)}} M_1 \cdots M_{u_F-1}$, and $\sigma = \rho_{\vec{j}^{(\ell)}}$. By use of the triangular inequality we also observe that

$$\begin{aligned} &D \left(M_{u_F-1} \cdots M_1 \rho_{\vec{j}^{(\ell)}} M_1 \cdots M_{u_F-1}, \rho_{\vec{j}^{(\ell)}} \right) \\ &\leq D \left(M_{u_F-1} \rho_{\vec{j}^{(\ell)}} M_{u_F-1}, \rho_{\vec{j}^{(\ell)}} \right) + D \left(M_{u_F-1} \cdots M_1 \rho_{\vec{j}^{(\ell)}} M_1 \cdots M_{u_F-1}, M_{u_F-1} \rho_{\vec{j}^{(\ell)}} M_{u_F-1} \right) \\ &\leq D \left(M_{u_F-1} \rho_{\vec{j}^{(\ell)}} M_{u_F-1}, \rho_{\vec{j}^{(\ell)}} \right) + D \left(M_{u_F-2} \cdots M_1 \rho_{\vec{j}^{(\ell)}} M_1 \cdots M_{u_F-2}, \rho_{\vec{j}^{(\ell)}} \right) \\ &\leq \sum_{u=1}^{u_F-1} D \left(M_u \rho_{\vec{j}^{(\ell)}} M_u, \rho_{\vec{j}^{(\ell)}} \right), \end{aligned} \quad (46)$$

where the second inequality follows from Lemma 4 while the third one by direct iteration of the previous passages. Replaced into Eq. (45) this finally yields

$$\langle p_{succ}(\ell) \rangle_{\mathcal{S}} \geq \left\langle \text{Tr} \left[M_{u_F}^2 \rho_{\vec{j}^{(\ell)}} \right] \right\rangle_{\mathcal{S}} - 2 \sum_{u=1}^{u_F-1} \left\langle D \left(M_u \rho_{\vec{j}^{(\ell)}} M_u, \rho_{\vec{j}^{(\ell)}} \right) \right\rangle_{\mathcal{S}}. \quad (47)$$

Assume now that Eq. (43) holds. Accordingly for all $u \in \{1, \dots, u_F\}$, $\ell = 1, \dots, N$ we have

$$\left\langle \text{Tr} \left[M_u^2 \rho_{\vec{j}^{(\ell)}} \right] \right\rangle_{\mathcal{S}} \geq 1 - \epsilon(n), \quad (48)$$

with $\epsilon(n)$ being a positive function which goes to zero faster than $1/n^2$. Then thanks to Lemma 2 we can write

$$\langle p_{succ}(\ell) \rangle_{\mathcal{S}} \geq 1 - \epsilon(n) - 2nR\sqrt{\epsilon(n)}, \quad (49)$$

which forces $\langle p_{succ}(\ell) \rangle_{\mathcal{S}}$ to converge to 1 as $n \rightarrow \infty$. This shows that Eq. (43) is indeed a sufficient condition for (42).

Part ii): To prove that Eq. (44) is a sufficient condition for (42) we invoke Lemma 1 with $E = M_u^2$, $\rho = \rho_{\vec{j}(\ell)}$ and $\sigma = P\rho_{\vec{j}(\ell)}P$ obtaining the following inequality

$$\left\langle \text{Tr} \left[M_u^2 \rho_{\vec{j}(\ell)} \right] \right\rangle_{\mathcal{S}} \geq \left\langle \text{Tr} \left[M_u^2 P \rho_{\vec{j}(\ell)} P \right] \right\rangle_{\mathcal{S}} - 2D \left(\rho_{\vec{j}(\ell)}, P \rho_{\vec{j}(\ell)} P \right). \quad (50)$$

From Eq. (16) we also know that for n sufficiently large and $\epsilon_1 = O(e^{-n})$ one has

$$\left\langle \text{Tr} \left[P \rho_{\vec{j}(\ell)} \right] \right\rangle_{\mathcal{S}} = \text{Tr} \left[P \left\langle \rho_{\vec{j}(\ell)} \right\rangle_{\mathcal{S}} \right] = \text{Tr} \left[P \rho^{\otimes n} \right] \geq 1 - \epsilon_1, \quad (51)$$

where we used the fact that the average over \mathcal{S} of the ℓ -th codeword corresponds to the average with respect to the joint probability (8) of $\rho_{\vec{j}}$, i.e.

$$\left\langle \rho_{\vec{j}(\ell)} \right\rangle_{\mathcal{S}} = \sum_{\vec{j}} p_{\vec{j}} \rho_{\vec{j}} = \rho^{\otimes n}. \quad (52)$$

Accordingly via Lemma 2 we can conclude that

$$\left\langle D \left(\rho_{\vec{j}(\ell)}, P \rho_{\vec{j}(\ell)} P \right) \right\rangle_{\mathcal{S}} \leq \sqrt{\epsilon_1}, \quad (53)$$

which inserted in Eq. (50) finally yields

$$\left\langle \text{Tr} \left[M_u^2 \rho_{\vec{j}(\ell)} \right] \right\rangle_{\mathcal{S}} \geq \left\langle \text{Tr} \left[M_u^2 P \rho_{\vec{j}(\ell)} P \right] \right\rangle_{\mathcal{S}} - 2\sqrt{\epsilon_1}, \quad (54)$$

which thanks to (44) implies the inequality (43) and hence, via part i) of the theorem, Eq. (42). \square

Thanks to Theorem 1 we can now prove that bisection protocols allows one to attain the Holevo bound, by showing that, for all rates R respecting (6), it is possible to identify operators $\{N_{k_1, \dots, k_u}^{(u)}\}_{u \in \{1, \dots, u_F = nR\}; k_1, \dots, k_{u-1} \in \{0, 1\}}$ fulfilling Eq. (44). Ideally one way of building the POVMs $\mathcal{M}_{k_1, k_2, \dots, k_{u-1}}^{(u)}$ which define the bisection decoding procedure, would be to identify its elements $N_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)}$, $N_{k_1, k_2, \dots, k_{u-1}, 1}^{(u)}$ with the projectors on the subspaces spanned by the codewords of sets $\mathbf{C}_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)}$ and $\mathbf{C}_{k_1, k_2, \dots, k_{u-1}, 1}^{(u)}$ respectively. This is not possible however due to the fact that such spaces are in general not orthogonal, though we expect typical subspaces of different codewords of the source to be disjoint in the long n limit: some kind of regularization is hence necessary. In the following we shall present three alternative, yet asymptotically equivalent, ways to realize this: the first makes use of orthogonal projections on subspaces identified by treating asymmetrically the set $\mathbf{C}_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)}$ and $\mathbf{C}_{k_1, k_2, \dots, k_{u-1}, 1}^{(u)}$, the second is based on the PGM construction, and finally the third makes use of the POVM elements of the sequential protocol of Refs. 13–15.

1. Method 1: orthogonal projections

Consider the set $\mathbf{C}_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)}$. For each one of its codewords $\rho_{\vec{j}^{(\ell)}}$ we can associate a typical subspace $\mathcal{H}_{typ}^{\vec{j}^{(\ell)}}$ and a corresponding projector $P_{\vec{j}^{(\ell)}}$ along the lines detailed in Sec. III A. Next we construct the subspace $\mathcal{H}_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)}$ spanned by the vectors which can be written as a direct sum of the elements of the $\mathcal{H}_{typ}^{\vec{j}^{(\ell)}}$ of $\mathbf{C}_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)}$, i.e.

$$\mathcal{H}_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)} := \bigoplus_{\ell \in \mathbf{C}_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)}} \mathcal{H}_{typ}^{\vec{j}^{(\ell)}} \quad (55)$$

where the sum is performed over the ℓ s whose corresponding vector $\rho_{\vec{j}^{(\ell)}}$ belongs to the set $\mathbf{C}_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)}$. By construction it follows that each one of the $\mathcal{H}_{typ}^{\vec{j}^{(\ell)}}$ associated to $\mathbf{C}_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)}$ are proper subspaces of $\mathcal{H}_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)}$. Accordingly, indicating with $P_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)}$ the projector on $\mathcal{H}_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)}$ we have that

$$P_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)} \geq P_{\vec{j}^{(\ell)}}, \quad (56)$$

for all $\ell \in \mathbf{C}_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)}$. Also due to the partial overlapping among the $\mathcal{H}_{typ}^{\vec{j}^{(\ell)}}$ of $\mathbf{C}_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)}$ the sum of the associated $P_{\vec{j}^{(\ell)}}$ s will in general be larger than $P_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)}$, i.e.

$$P_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)} \leq \sum_{\ell \in \mathbf{C}_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)}} P_{\vec{j}^{(\ell)}}. \quad (57)$$

We define the orthogonal projections method for the bisection POVM by identifying $N_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)}$ with $P_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)}$ and $N_{k_1, k_2, \dots, k_{u-1}, 1}^{(u)}$ with its complementary counterpart, i.e.

$$N_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)} := P_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)}, \quad (58)$$

$$N_{k_1, k_2, \dots, k_{u-1}, 1}^{(u)} := Q_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)} = \mathbf{1} - P_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)}. \quad (59)$$

A couple of remarks are mandatory:

- i) notice that $N_{k_1, k_2, \dots, k_{u-1}, 1}^{(u)}$ does not coincide with the projector $P_{k_1, k_2, \dots, k_{u-1}, 1}^{(u)}$ on the subspace $\mathcal{H}_{k_1, k_2, \dots, k_{u-1}, 1}^{(u)}$ formed by the direct sum of the typical subspaces $\mathcal{H}_{typ}^{\vec{j}^{(\ell)}}$ associated with $\mathbf{C}_{k_1, k_2, \dots, k_{u-1}, 1}^{(u)}$. Notice also that, due to the partial overlapping of the typical subspaces of different codewords, in general we can neither establish an inequality similar to Eq. (56) which links $N_{k_1, k_2, \dots, k_{u-1}, 1}^{(u)}$ and the $P_{\vec{j}^{(\ell)}}$ of $\mathcal{H}_{k_1, k_2, \dots, k_{u-1}, 1}^{(u)}$, nor fix an ordering between $N_{k_1, k_2, \dots, k_{u-1}, 1}^{(u)}$ and $P_{k_1, k_2, \dots, k_{u-1}, 1}^{(u)}$;

ii) by construction the scheme we are analyzing here does not include the possibility of the null event described in the previous section. Indeed in this case we have

$$N_{k_1, k_2, \dots, k_{u-1}, \text{null}}^{(u)} = \mathbf{1} - N_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)} - N_{k_1, k_2, \dots, k_{u-1}, 1}^{(u)} = 0. \quad (60)$$

The associated set POVM $\mathcal{M}_{k_1, k_2, \dots, k_{u-1}}^{(u)}$ is thus a projective measurement which admits only two possible outcomes, $k_u = 0$ and $k_u = 1$.

From Theorem 1 the asymptotic attainability of the Holevo bound with this procedure can be established by showing that Eq. (44) holds, i.e. explicitly

Lemma 6. *For all rates R satisfying the Holevo bound (6) the bisection scheme associated with operators $\{N_{k_1, \dots, k_u}^{(u)}\}$ defined as in Eqs. (58), (59) fullfils the sufficient condition (44).*

Proof. Consider first the case with $k_u = 0$. Given then a generic codeword $\rho_{\vec{j}^{(\ell)}}^{(u)}$ of $\mathbf{C}_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)}$ we can write

$$\begin{aligned} \left\langle \text{Tr} \left[N_{k_1^{(\ell)}, k_2^{(\ell)}, \dots, k_{u-1}^{(\ell)}, 0}^{(u)} \bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_{\mathcal{S}} &= \left\langle \text{Tr} \left[P_{k_1^{(\ell)}, k_2^{(\ell)}, \dots, k_{u-1}^{(\ell)}, 0}^{(u)} \bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_{\mathcal{S}} \geq \left\langle \text{Tr} \left[P_{\vec{j}^{(\ell)}} \bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_{\mathcal{S}} \\ &= \sum_{\vec{j}} p_{\vec{j}} \text{Tr} \left[P_{\vec{j}} \bar{\rho}_{\vec{j}} \right] \geq \sum_{\vec{j}} p_{\vec{j}} \left(\text{Tr} \left[P_{\vec{j}} \rho_{\vec{j}} \right] - 2D \left(\bar{\rho}_{\vec{j}}, \rho_{\vec{j}} \right) \right) \geq 1 - \epsilon_2 - 2\sqrt{\epsilon_1}, \end{aligned} \quad (61)$$

where for easy of notation we set $\bar{\rho}_{\vec{j}^{(\ell)}} := P \rho_{\vec{j}^{(\ell)}} P$ and where we used the fact that taking the average with respect to the statistical collection \mathcal{S} of $\text{Tr} \left[P_{\vec{j}^{(\ell)}} \bar{\rho}_{\vec{j}^{(\ell)}} \right]$ is equivalent to taking the average of $\text{Tr} \left[P_{\vec{j}} \bar{\rho}_{\vec{j}} \right]$ with respect to $p_{\vec{j}}$, i.e.

$$\left\langle \text{Tr} \left[P_{\vec{j}^{(\ell)}} \bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_{\mathcal{S}} = \sum_{\vec{j}} p_{\vec{j}} \text{Tr} \left[P_{\vec{j}} \bar{\rho}_{\vec{j}} \right]. \quad (62)$$

The first inequality of Eq. (61) follows from Eq. (56); the second inequality follows instead from applying Lemma 1 with $E = P_{\vec{j}}$, $\rho = \bar{\rho}_{\vec{j}}$ and $\sigma = \rho_{\vec{j}}$; while finally the third inequality follows both from the high probability of projecting codeword $\rho_{\vec{j}}$ on its conditionally typical subspace (24) and from the same concept for the average codeword, together with Lemma 2 (as in (51,53)), the parameters ϵ_1 and ϵ_2 being both exponentially small in n to guarantee the limit property (19). Equation (61) proves hence that Eq. (44) applies at least for the sets $\mathbf{C}_{k_1, k_2, \dots, k_u}^{(u)}$ with $k_u = 0$.

Take next $k_u = 1$ and a generic codeword $\rho_{\vec{j}^{(\ell)}}^{(u)}$ of $\mathbf{C}_{k_1, k_2, \dots, k_{u-1}, 1}^{(u)}$. In this case we have

$$\begin{aligned} \left\langle \text{Tr} \left[N_{k_1^{(\ell)}, k_2^{(\ell)}, \dots, k_{u-1}^{(\ell)}, 1}^{(u)} \bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_{\mathcal{S}} &= \left\langle \text{Tr} \left[\bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_{\mathcal{S}} - \left\langle \text{Tr} \left[P_{k_1^{(\ell)}, k_2^{(\ell)}, \dots, k_{u-1}^{(\ell)}, 0}^{(u)} \bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_{\mathcal{S}} \\ &\geq \left\langle \text{Tr} \left[\bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_{\mathcal{S}} - \sum_{\ell' \neq \ell} \left\langle \text{Tr} \left[P_{\vec{j}^{(\ell')}} \bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_{\mathcal{S}} \end{aligned} \quad (63)$$

where the inequality follows from (57) plus adding all the remaining terms $P_{\vec{j}^{(\ell')}}$ associated with codewords having $\ell' \neq \ell$. Observe then that from Eq. (16) we have

$$\left\langle \text{Tr} \left[\bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_{\mathcal{S}} = \sum_{\vec{j}} p_{\vec{j}} \text{Tr} \left[P \rho_{\vec{j}} \right] = \text{Tr} \left[P \rho^{\otimes n} \right] \geq 1 - \epsilon_1, \quad (64)$$

with ϵ_1 being an exponentially small function of n . Furthermore for each term of the sum on the RHS of Eq. (66) we have

$$\begin{aligned} \left\langle \text{Tr} \left[P_{\vec{j}^{(\ell')}} \bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_{\mathcal{S}} &= \sum_{\vec{j}, \vec{j}'} p_{\vec{j}} p_{\vec{j}'} \text{Tr} \left[\bar{\rho}_{\vec{j}} P_{\vec{j}'} \right] = \sum_{\vec{j}'} p_{\vec{j}'} \text{Tr} \left[\overline{\rho^{\otimes n}} P_{\vec{j}'} \right] \leq \|\overline{\rho^{\otimes n}}\|_{\infty} \sum_{\vec{j}'} p_{\vec{j}'} \text{Tr} \left[P_{\vec{j}'} \right] \\ &\leq 2^{-n[S(\rho) - \delta]} 2^{n[\sum_j p_j S(\rho_j) + \delta]} = 2^{-n[\chi(\{p_j, \rho_j\}) - 2\delta]}, \end{aligned} \quad (65)$$

where the second inequality follows from typical subspaces' properties (18,25) and where $\chi(\{p_j, \rho_j\})$ is Holevo information (7) of the source \mathcal{E} . Replacing (64) and (65) into Eq. (63) we arrive hence to

$$\left\langle \text{Tr} \left[N_{k_1^{(\ell)}, k_2^{(\ell)}, \dots, k_{u-1}^{(\ell)}, 1}^{(u)} \bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_{\mathcal{S}} \geq 1 - \epsilon_1 - 2^{nR} 2^{-n[\chi(\{p_j, \rho_j\}) - 2\delta]}, \quad (66)$$

which shows that as long as the rate R respects the Holevo bound (6), i.e.

$$R < \chi(\{p_j, \rho_j\}) - 2\delta, \quad (67)$$

for some $\delta > 0$, Eq. (44) applies also for the sets $\mathbf{C}_{k_1, k_2, \dots, k_u}^{(u)}$ with $k_u = 1$. \square

The inequalities (61) and (66) prove that under the constraint (67) the proposed implementation of the bisection decoding scheme asymptotically attains the Holevo bound, yielding an average error probability which converges to zero in the limit of $n \rightarrow \infty$.

2. Method 2: via PGM detections

An alternative way to implement the bisection protocol is substituting the sequential set measurement N with one inspired by the Pretty Good Measurement (PGM), first introduced to demonstrate the achievability of the Holevo bound.

For each set $\mathbf{C}_{k_1, \dots, k_u}^{(u)}$ define the positive operator

$$S_{k_1, \dots, k_u}^{(u)} = \sum_{\ell' \in \mathbf{C}_{k_1, \dots, k_u}^{(u)}} P_{\vec{j}(\ell')}, \quad (68)$$

i.e. the sum of projectors of all the codewords in that set. From the non-orthogonality of projectors and the completeness property (37) it follows

$$S_{k_1, \dots, k_{u-1}}^{(u-1)} = S_{k_1, \dots, k_{u-1}, 0}^{(u)} + S_{k_1, \dots, k_{u-1}, 1}^{(u)} \geq \mathbf{1}. \quad (69)$$

Thus we can build the u -th measurement to decide whether the word belongs to $\mathbf{C}_{k_1, \dots, k_{u-1}, 0}^{(u)}$ or $\mathbf{C}_{k_1, \dots, k_{u-1}, 1}^{(u)}$ by using the sum operators for these two sets, renormalized by the sum operator for $\mathbf{C}_{k_1, \dots, k_{u-1}}^{(u-1)}$, which contains both of them at the previous step:

$$N_{k_1, \dots, k_u}^{(u)} = \left[S_{k_1, \dots, k_{u-1}}^{(u-1)} \right]^{-1/2} S_{k_1, \dots, k_u}^{(u)} \left[S_{k_1, \dots, k_{u-1}}^{(u-1)} \right]^{-1/2}, \quad (70)$$

where the inverse $\left[S_{k_1, \dots, k_{u-1}}^{(u-1)} \right]^{-1/2}$ is meant to be computed only on the support of $S_{k_1, \dots, k_{u-1}}^{(u-1)}$ (otherwise the operator is assumed to be null). In this way we obtain a proper set POVM, since the renormalization allows us to take into account the intersections between typical subspaces of different codewords, i.e.

$$0 \leq N_{k_1, \dots, k_u}^{(u)} \leq \sum_{k \in \{0,1\}} N_{k_1, \dots, k}^{(u)} = \left[S_{k_1, \dots, k_{u-1}}^{(u-1)} \right]^{-1/2} \left[S_{k_1, \dots, 0}^{(u)} + S_{k_1, \dots, 1}^{(u)} \right] \left[S_{k_1, \dots, k_{u-1}}^{(u-1)} \right]^{-1/2} \leq \mathbf{1}.$$

As in the case discussed previously we can now show that for all R fulfilling the Holevo bound (6) the operators (70) satisfy the sufficient condition Eq. (44) of Theorem 1.

Lemma 7. *For all rates R satisfying the Holevo bound (6) the bisection scheme associated with operators $\{N_{k_1, \dots, k_u}^{(u)}\}$ defined as in Eq. (70) fulfils the sufficient condition (44).*

Proof. Observe that

$$\begin{aligned} \left\langle \text{Tr} \left[N_{k_1^{(\ell)}, \dots, k_u^{(\ell)}}^{(u)} \bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_S &\geq \left\langle \text{Tr} \left[\left[S_{k_1^{(\ell)}, \dots, k_{u-1}^{(\ell)}}^{(u-1)} \right]^{-1/2} P_{\vec{j}^{(\ell)}} \left[S_{k_1^{(\ell)}, \dots, k_{u-1}^{(\ell)}}^{(u-1)} \right]^{-1/2} \bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_S \\ &= \left\langle \text{Tr} \left[\Lambda_\ell \bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_S = \langle p_{\text{succ}}^{(u-1)}(\ell) \rangle_S, \end{aligned} \quad (71)$$

where the latter is the average success probability of recovering the ℓ -th codeword from the set $\mathbf{C}_{k_1^{(\ell)}, \dots, k_{u-1}^{(\ell)}}^{(u-1)}$ while using a PGM strategy and Λ_ℓ is the corresponding POVM element.

Accordingly we can bound each of the terms on the RHS of Eq. (47) by exploiting the efficiency of the PGM protocol. Specifically, we employ the Hayashi-Nagaoka inequality¹¹

$$\mathbf{1} - \Lambda_\ell \leq 2Q_{\vec{j}^{(\ell)}} + 4 \sum_{\ell' \neq \ell} P_{\vec{j}^{(\ell')}} \quad (72)$$

to write the average success probability as

$$\langle p_{succ}^{(u-1)}(\ell) \rangle_S \geq \left\langle \text{Tr} \left[\bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_S - 2 \left\langle \text{Tr} \left[Q_{\vec{j}^{(\ell)}} \bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_S - 4 \sum_{\ell' \neq \ell} \left\langle P_{\vec{j}^{(\ell')}} \bar{\rho}_{\vec{j}^{(\ell)}} \right\rangle_S \quad (73)$$

$$\geq 1 - \epsilon_1 - 2(\epsilon_2 + 2\sqrt{\epsilon_1}) - 4 \cdot 2^{nR} \cdot 2^{-n[\chi(\{p_j, \rho_j\}) - 2\delta]}, \quad (74)$$

where the last inequality follows from Eqs. (51) and (65) and the fact that

$$\begin{aligned} \left\langle \text{Tr} \left[\bar{\rho}_{\vec{j}^{(\ell)}} Q_{\vec{j}^{(\ell)}} \right] \right\rangle_S &= \left\langle \text{Tr} \left[\bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_S - \left\langle \text{Tr} \left[P_{\vec{j}^{(\ell)}} \bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_S \\ &\leq 1 - \left\langle \text{Tr} \left[P_{\vec{j}^{(\ell)}} \bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_S = 1 - \sum_{\vec{j}} p_{\vec{j}} \text{Tr} \left[P_{\vec{j}} \bar{\rho}_{\vec{j}} \right] \\ &\leq 1 - \sum_{\vec{j}} p_{\vec{j}} \left(\text{Tr} \left[P_{\vec{j}} \rho_{\vec{j}} \right] - 2D \left(\bar{\rho}_{\vec{j}}, \rho_{\vec{j}} \right) \right) \leq \epsilon_2 + 2\sqrt{\epsilon_1}, \end{aligned} \quad (75)$$

which is derived as in Eq.(61). Similarly to what we observed in Eq. (66) it then follows that if the rate R fulfills the constraint (67) for some $\delta > 0$, then for n sufficiently large one has that, for all u and ℓ ,

$$\left\langle \text{Tr} \left[N_{k_1^{(\ell)}, \dots, k_u^{(\ell)}}^{(u)} \bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_S \geq \left\langle \text{Tr} \left[\Lambda_\ell \bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_S \geq 1 - \epsilon_3, \quad (76)$$

with $\epsilon_3 = O(e^{-n})$ being exponentially small in n and fulfilling the condition (19) showing hence that (44) is satisfied by the selected operators. \square

3. Method 3: via sequential POVM

Another way to regularize set-projection operators necessary to implement the bisection scheme, is to make use of the sequential protocol for that set, but without gaining knowledge about the result of this subroutine. Accordingly the regularized set-projection operators will be implemented as a black box, applying the sequential decoding scheme to the set of codewords which appear inside that set, taking also into account failure in projecting on the typical subspace of previous codewords, in the code ordering chosen by Bob, see Fig. 2. The resulting setting is clearly redundant as the vast majority of information

gathered via the sequential decoding is simply neglected in the process. Also, the same procedure is iterated every time a new bit of the bijective encoding has to be acquired, increasing hence the chances of deteriorating the transmitted codeword. Still, as we shall see in the following, the scheme is efficient enough to allow for the saturation of the Holevo bound.

In order to formalize this construction, for each quantum codeword $\rho_{\vec{j}^{(\ell)}} \in \mathbf{C}$, we write its corresponding element of the sequential POVM^{13–15} as

$$\begin{aligned} E_1 &= P_{\vec{j}^{(1)}} , \\ E_\ell &= Q_{\vec{j}^{(1)}} \dots Q_{\vec{j}^{(\ell-1)}} P_{\vec{j}^{(\ell)}} Q_{\vec{j}^{(\ell-1)}} \dots Q_{\vec{j}^{(1)}} , \quad \ell \geq 2 , \end{aligned} \quad (77)$$

where $P_{\vec{j}}$ is the projector on the typical subspace of codeword state $\rho_{\vec{j}}$ and $Q_{\vec{j}} = \mathbf{1} - P_{\vec{j}}$ its complementary. We remind that by construction these operators fulfill the proper normalization condition,

$$0 \leq E_\ell \leq \mathbf{1} , \quad (78)$$

$$0 \leq \sum_{\ell=1}^N E_\ell = \mathbf{1} - E_0 \leq \mathbf{1} , \quad (79)$$

and that, given a density matrix $\rho_{\vec{j}} \in \mathbf{C}$, the probability of recovering the codeword $\vec{j}^{(\ell)}$ is given by

$$P_{seq}(\vec{j}^{(\ell)} | \rho_{\vec{j}}) = \text{Tr}[E_\ell \rho_{\vec{j}}] . \quad (80)$$

Using this expression we can hence estimate the probability that $\rho_{\vec{j}^{(\ell)}}$ belongs to the set $\mathbf{C}_{k_1, k_2, \dots, k_u}^{(u)}$ by simply summing the above expression over all $\vec{j}^{(\ell)}$ belonging to such set, i.e.

$$P(\rho_{\vec{j}^{(\ell)}} \in \mathbf{C}_{k_1, k_2, \dots, k_u}^{(u)}) = \sum_{\ell \in \mathbf{C}_{k_1, k_2, \dots, k_u}^{(u)}} P_{seq}(\vec{j}^{(\ell)} | \rho_{\vec{j}}) = \text{Tr}[N_{k_1, k_2, \dots, k_u}^{(u)} \rho_{\vec{j}}] , \quad (81)$$

where the sum is performed over the ℓ s whose corresponding vector $\rho_{\vec{j}^{(\ell)}}$ belongs to the set $\mathbf{C}_{k_1, k_2, \dots, k_u}^{(u)}$, and where

$$N_{k_1, k_2, \dots, k_u}^{(u)} = \sum_{\ell \in \mathbf{C}_{k_1, k_2, \dots, k_u}^{(u)}} E_\ell , \quad (82)$$

is the set-sequential-measurement associated with the set $\mathbf{C}_{k_1, k_2, \dots, k_u}^{(u)}$ induced by the sequential decoding POVM. Since for all $u \in \{1, \dots, u_F\}$ and for all k_1, k_2, \dots, k_{u-1} , the sets

$\mathbf{C}_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)}$ and $\mathbf{C}_{k_1, k_2, \dots, k_{u-1}, 1}^{(u)}$ are not overlapping (see e.g. Eq. (36)) we have that

$$0 \leq N_{k_1, k_2, \dots, k_u}^{(u)} \leq \sum_{k \in \{0,1\}} N_{k_1, k_2, \dots, k}^{(u)} \leq \sum_{\ell=1}^N E_\ell \leq \mathbf{1}, \quad (83)$$

which guarantees that the operators $N_{k_1, k_2, \dots, k_{u-1}, 0}^{(u)}$, $N_{k_1, k_2, \dots, k_{u-1}, 1}^{(u)}$, and $N_{k_1, k_2, \dots, k_{u-1}, \text{null}}^{(u)} = \mathbf{1} - \sum_{k \in \{0,1\}} N_{k_1, k_2, \dots, k}^{(u)}$ form a properly normalized POVM.

Lemma 8. *For all rates R satisfying the Holevo bound (6) the bisection scheme associated with operators $\{N_{k_1, \dots, k_u}^{(u)}\}$ defined as in Eq. (82) fullfils the sufficient condition (44).*

Proof. First observe that each operator $N_{k_1^{(\ell)}, \dots, k_u^{(\ell)}}^{(u)}$ is the sum of a certain number of sequential POVM elements, always containing the element E_ℓ corresponding to the right codeword. Since all the operators in the sum are positive we can state

$$\left\langle \text{Tr} \left[N_{k_1^{(\ell)}, \dots, k_u^{(\ell)}}^{(u)} \bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_{\mathcal{S}} = \sum_{\ell' \in \mathbf{C}_{k_1^{(\ell)}, \dots, k_u^{(\ell)}}^{(u)}} \left\langle \text{Tr} \left[E_{\ell'} \bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_{\mathcal{S}} \geq \left\langle \text{Tr} \left[E_\ell \bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_{\mathcal{S}}, \quad (84)$$

where in the last term we recognize the average success probability (80) of the sequential protocol computed on the subnormalized version $\bar{\rho}_{\vec{j}^{(\ell)}}$ of the ℓ -th codeword. Accordingly we can bound each of the terms on the RHS of Eq. (47) by exploiting the efficiency of the sequential protocol. Specifically by applying Sen's Lemma 5 and using the concavity of the square root function we can write:

$$\left\langle \text{Tr} \left[E_\ell \bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_{\mathcal{S}} = \left\langle \text{Tr} \left[P_{\vec{j}^{(\ell)}} Q_{\vec{j}^{(\ell-1)}} \dots Q_{\vec{j}^{(1)}} \bar{\rho}_{\vec{j}^{(\ell)}} Q_{\vec{j}^{(1)}} \dots Q_{\vec{j}^{(\ell-1)}} P_{\vec{j}^{(\ell)}} \right] \right\rangle_{\mathcal{S}} \quad (85)$$

$$\begin{aligned} &\geq \left\langle \text{Tr} \left[\bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_{\mathcal{S}} - 2 \left\langle \sqrt{\text{Tr} \left[\bar{\rho}_{\vec{j}^{(\ell)}} Q_{\vec{j}^{(\ell)}} \right] + \sum_{\ell' \neq \ell} \text{Tr} \left[\bar{\rho}_{\vec{j}^{(\ell)}} P_{\vec{j}^{(\ell')}} \right]} \right\rangle_{\mathcal{S}} \\ &\geq \left\langle \text{Tr} \left[\bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_{\mathcal{S}} - 2 \sqrt{\left\langle \text{Tr} \left[\bar{\rho}_{\vec{j}^{(\ell)}} Q_{\vec{j}^{(\ell)}} \right] \right\rangle_{\mathcal{S}} + \sum_{\ell' \neq \ell} \left\langle \text{Tr} \left[\bar{\rho}_{\vec{j}^{(\ell)}} P_{\vec{j}^{(\ell')}} \right] \right\rangle_{\mathcal{S}}}, \end{aligned} \quad (86)$$

having added under square root all the terms $P_{\vec{j}^{(\ell')}}$ with $\ell' > \ell$. The term outside the square-root can be treated as in (51). For the first term under square-root simply apply Eq. (75). For each term of the sum under square-root we can instead use the inequality (65). Therefore we can write

$$\left\langle \text{Tr} \left[E_\ell \bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_{\mathcal{S}} \geq 1 - \epsilon_1 - 2 \sqrt{\epsilon_2 + 2 \sqrt{\epsilon_1} + 2^{nR} \cdot 2^{-n[\chi(\{p_j, \rho_j\}) - 2\delta]}}, \quad (87)$$

which, via Eq. (84) implies again that for rates R fulfilling Eq. (67) for some $\delta > 0$, then for n sufficiently large one has that, for all u and ℓ ,

$$\left\langle \text{Tr} \left[N_{k_1^{(\ell)}, \dots, k_u^{(\ell)}}^{(u)} \bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_{\mathcal{S}} \geq \left\langle \text{Tr} \left[E_{\ell} \bar{\rho}_{\vec{j}^{(\ell)}} \right] \right\rangle_{\mathcal{S}} \geq 1 - \epsilon_3, \quad (88)$$

with $\epsilon_3 = O(e^{-n})$ being exponentially small in n and fulfilling the condition (19). This proves (44) and hence the asymptotic achievability of the Holevo bound with the bisection protocol defined by operators (82). \square

V. CONCLUSIONS

In this article we computed an upper bound for the average error probability (over all codewords in a code and over all possible codes) of the bisection decoding scheme. The bound is shown to approach zero exponentially fast with the codewords' length, for any output ensemble \mathcal{E} whose size is strictly less than $2^{\chi(\mathcal{E})}$. Thus we provided a new proof of the attainability of the Holevo bound for classical communication through a quantum channel for a class of decoding schemes based on the bisection method, whose complexity scales as the logarithm of the codewords' length. An advantage of this protocol is the possibility of gaining a bit of information at each step of the procedure, unlike the sequential decoding, which gives either full or null information about the codeword at each step. This is particularly powerful in the case of failure at a certain step of the protocol, allowing the receiver to at least make use of the previous steps for a partial identification of the message. Note also that there is a certain degree of freedom in the implementation of the specific sets' "yes-no" measurements, which form a complete POVM at each step, independently of the rest of the protocol, as long as their average error probability approaches zero faster than n^{-2} (e.g. exponentially decaying) as the codewords' length n grows, for all sources respecting the Holevo bound. This fact has been shown by providing three different POVMs which satisfy the bound, employing projectors on typical subspaces and renormalizing for their non-orthogonality. Unfortunately, as in the case of polar coding, this general requirements on the "yes-no" set measurements do not allow one to evaluate their implementation complexity, thus leaving open the problem of determining a structured device for the efficient detection of long codewords. Eventually we stress the importance of the Chernoff bound to provide an exponential scaling to the small quantities used in describing the typical subspaces'

properties, which in turn allows the convergence of the decoding scheme.

Appendix A: The law of large numbers via Chernoff bound

In this appendix we compute an exponential bound for the law of large numbers, which guarantees the convergence of the error probability of our protocol to zero. Indeed consider the small quantities ϵ_1, ϵ_2 which appear in Sec. III A. These quantities describe the high probability of finding respectively the average state $\rho^{\otimes n}$ and the codeword states $\rho_{\vec{j}}$ in their typical subspaces, identified by the projectors P and $P_{\vec{j}}$. This is why they are connected, through the classical typical subspaces, to the law of large numbers.

Consider for example the average state ρ of the source. We can easily prove that the probability of n copies of the quantum state ρ are in its δ -typical subspace, $\text{Tr}[P\rho^{\otimes n}]$, is equivalent to the probability of a random sample sequence \vec{x} of the corresponding classical source being in the classical δ -typical subspace, $Pr(\vec{x} \in T_\delta^n)$:

$$\text{Tr}[P\rho^{\otimes n}] = \text{Tr}\left[\sum_{\vec{x} \in T_\delta^n} |e_{\vec{x}}\rangle\langle e_{\vec{x}}| \sum_{\vec{x}'} q_{\vec{x}'} |e_{\vec{x}'}\rangle\langle e_{\vec{x}'}|\right] \quad (\text{A1})$$

$$= \sum_{\vec{x} \in T_\delta^n} \sum_{\vec{x}'} |\langle e_{\vec{x}} | e_{\vec{x}'} \rangle|^2 q_{\vec{x}'} \quad (\text{A2})$$

$$= \sum_{\vec{x} \in T_\delta^n} q_{\vec{x}} = Pr(\vec{x} \in T_\delta^n). \quad (\text{A3})$$

A similar result is obtained for each codeword state $\rho_{\vec{j}}$, namely

$$\text{Tr}[P_{\vec{j}}\rho_{\vec{j}}] = Pr(\vec{y} \in T_\delta^{\vec{j}}). \quad (\text{A4})$$

These probabilities can be bounded from above with the help of the law of large numbers. Consider for example the average typical subspace and choose the random variable Z , taking values $z = -\log_2 q_x$. We also choose the same probability distribution both for X and Z , i.e. $q_z \equiv q_x$. Then the law of large numbers states that, for any $\delta > 0$, the probability that the average of Z over n extractions,

$$\frac{1}{n} \sum_{i=1}^n z_i = \bar{H}(\vec{x}), \quad (\text{A5})$$

i.e. the sum of n i.i.d. random variables, differs from its expected value,

$$\sum_{i=1}^n q_{z_i} z_i = H(X), \quad (\text{A6})$$

for more than δ is lower than a small and positive quantity $1 \gg \epsilon > 0$, i.e.

$$Pr(\vec{x} \in T_\delta^n) = Pr(|\bar{H}(\vec{x}) - H(X)| \geq \delta) \leq \epsilon. \quad (\text{A7})$$

In usual derivations of this result the Chebyshev inequality is exploited, which gives a scaling behaviour $\epsilon \sim n^{-1}$. This is not sufficient for convergence of the error probability to 0 for long sequences $n \rightarrow \infty$ in (49). Recalling also that the Chebyshev bound gives a dependence on the variance of the distribution, it is clear that such a scaling is a rough estimate, since the law of large numbers is known to be valid also for infinite-variance distributions. We therefore use the Chernoff bound to obtain a faster, indeed exponential, convergence.

Consider first the Markov inequality, valid for any nonnegative random variable $t > 0$ and $\delta > 0$:

$$Pr(t \geq \delta) = \sum_{t \geq \delta} p_t \leq \sum_{t \geq \delta} p_t \frac{t}{\delta} \quad (\text{A8})$$

$$\leq \frac{1}{\delta} \sum_t t p_t = \frac{\bar{t}}{\delta}, \quad (\text{A9})$$

where we used a bar sign to indicate the average over the probability distribution of the random variable. The first inequality follows from introducing terms which certainly are less than one, given the constraint on the sum. The second inequality follows from adding positive terms to the sum, since the random variable is positive. We now choose $t = e^{sw}$, with w a new random variable[?] and $\delta = e^{sA}$, without loss of generality. The Markov inequality then reads

$$Pr(e^{sw} \geq e^{sA}) \leq e^{-sA} g_w(s) \quad (\text{A10})$$

for any s, A , where we called $g_w(s) = \overline{\exp(sw)}$ the moment generating function of the random variable w , i.e.

$$\overline{w^n} = \left. \frac{d^n g_w(s)}{ds^n} \right|_{s=0}. \quad (\text{A11})$$

Now observe that the above inequality between exponentials has two different meanings depending on the sign of s , implying both

$$Pr(w \geq A) \leq e^{-sA} g_w(s) \quad s > 0 \quad (\text{A12})$$

$$Pr(w \leq A) \leq e^{-sA} g_w(s) \quad s < 0. \quad (\text{A13})$$

These two relations give bounds on the tails of the w probability distribution. In order to evaluate how tight such bounds are, we consider the specific case of w being the sum of n

i.i.d random variables x_i , implying for the moment generating function

$$g_w(s) = \exp \left(\overline{s \sum_{i=1}^n x_i} \right) = \prod_{i=1}^n \overline{e^{sx_i}} = (g_x(s))^n, \quad (\text{A14})$$

and take $A = na$, without loss of generality. The previous inequalities become

$$Pr \left(\frac{1}{n} \sum_{i=1}^n x_i \geq a \right) \leq \exp [-n (sa - \ln g_x(s))] \quad s > 0 \quad (\text{A15})$$

$$Pr \left(\frac{1}{n} \sum_{i=1}^n x_i \leq a \right) \leq \exp [-n (sa - \ln g_x(s))] \quad s < 0. \quad (\text{A16})$$

We now need to evaluate the behaviour of the coefficient function in the exponential:

$$h(s) = sa - \ln g_x(s). \quad (\text{A17})$$

Consider first some properties of $\mu_x(s) = \ln g_x(s)$, following from the nature of the moment generating function:

- $\mu_x(s=0) = 0$, since $g_x(s=0) = 1$;
- $\mu'_x(s=0) = g'_x(s=0)/g_x(s=0) = \bar{x}$, since $g_x(s=0) = \bar{x}$;
- it is convex

$$\mu''_x(s) = \frac{g''_x(s)}{g_x(s)} - \left(\frac{g'_x(s)}{g_x(s)} \right)^2 \quad (\text{A18})$$

$$= \langle x^2 \rangle_e - \langle x \rangle_e^2 = \langle (x - \langle x \rangle_e)^2 \rangle_e \geq 0 \quad \forall s, \quad (\text{A19})$$

where we have indicated with

$$\langle f(x) \rangle_e = \frac{\overline{f(x)e^{sx}}}{e^{sx}} \quad (\text{A20})$$

the probability average with weight e^{sx} .

From the previous properties, it follows that the slope of the function, starting at \bar{x} at the origin, increases for $s > 0$ and decreases for $s < 0$. Expanding $\mu_x(s)$ for small s at second order, we have for the coefficient function

$$h(s) \simeq (a - \bar{x})s - \frac{s^2}{2} \mu''_x(0). \quad (\text{A21})$$

This approximate function (for small s) is zero at

$$s^* = \frac{2(a - \bar{x})}{\mu_x''(0)}. \quad (\text{A22})$$

Consider now the $s > 0$ inequality (A15). If $a > \bar{x}$, then the zero s^* is positive and inside the range of validity of the inequality. Thus $h(s) > 0$ for all $s < s^*$ in the range: the first inequality has a tight bound. Vice versa if $a < \bar{x}$, the zero s^* is negative and $h(s) < 0$ in the whole range of validity of the first inequality, making it useless.

The situation is reversed when considering the $s < 0$ inequality (A16). In this case we need $s^* < 0$, i.e. $a < \bar{x}$, and for any $s > s^*$ in the range the coefficient function will be positive again, providing a tight bound for the second inequality. By calling

$$h_p = \sup_{s>0} h(s), \quad h_m = \sup_{s<0} h(s), \quad h_p, h_m > 0, \quad (\text{A23})$$

the supremum of $h(s)$ in each region, we can thus rewrite the inequalities as tight bounds, taking respectively $a = \bar{x} + \delta > \bar{x}$ in the first inequality and $a = \bar{x} - \delta < \bar{x}$, in the second one, for any $\delta > 0$:

$$Pr \left(\frac{1}{n} \sum_{i=1}^n x_i - \bar{x} \geq \delta \right) \leq e^{-nh_p} \quad (\text{A24})$$

$$Pr \left(\frac{1}{n} \sum_{i=1}^n x_i - \bar{x} \leq -\delta \right) \leq e^{-nh_m}. \quad (\text{A25})$$

Eventually we sum the previous inequalities to obtain the law of large numbers with exponentially decreasing tails

$$Pr \left(\left| \frac{1}{n} \sum_{i=1}^n x_i - \bar{x} \right| \geq \delta \right) \leq e^{-nh_p} + e^{-nh_m} = O(e^{-n}) = \epsilon. \quad (\text{A26})$$

Observe that the small quantity $\epsilon > 0$ obtained in this way, exponentially decreasing with increasing n , also depends on the difference parameter δ that we chose, as of course is to be expected. Indeed this dependence is implicit in the definition of h_p, h_m : by choosing δ , we set different values of a (for both the $s > 0$ and $s < 0$ cases) and this in turn varies the point s^* (A22), i.e. the range of values of s ($s < s^*$ or $s > s^*$) which can be chosen to maximize the coefficient functions. In particular, since the expression (A21) is a small- s expansion, we do not know what the absolute supremum of $h(s)$ is and where it is located⁷. Thus by varying the range of s accessible through the tuning of δ , we may happen to exclude this

and other local supremum points, resulting in (possibly discontinuously) varying values of h_p, h_m .

In any case, for our purpose we need only the existence of a range of values s , both above and below zero and depending on δ , where the coefficient function $h(s)$ is positive, and this is guaranteed by the properties of the $\mu_x(s)$ function, respectively when $a > \bar{x}$ for positive s and when $a < \bar{x}$ for negative s .

Appendix B: Proofs of Lemmas

We give here the proofs of the remaining Lemmas of Section III B.

Proof of Lemma 3. For a hermitian operator we can always write $\omega = A - B$, where A, B are positive matrices with disjoint supports, representing ω respectively in the positive and negative part of its support. Consider then the operator $\bar{\Lambda} = \Pi_A - \Pi_B$, with Π_A and Π_B being projectors respectively on the support of A and of B . For this operator we can clearly state that $-\mathbf{1} \leq \bar{\Lambda} \leq \mathbf{1}$, i.e. for all vectors $|v\rangle$ we have

$$\langle v | (\bar{\Lambda} - \mathbf{1}) | v \rangle \leq 0 \quad (\text{B1})$$

$$\langle v | (\bar{\Lambda} + \mathbf{1}) | v \rangle \geq 0. \quad (\text{B2})$$

By construction we obtain thus an operator which saturates the bound (30):

$$\text{Tr} [\bar{\Lambda} \omega] = \text{Tr} [(\Pi_A - \Pi_B) A] - \text{Tr} [(\Pi_A - \Pi_B) B] \quad (\text{B3})$$

$$= \text{Tr} [A] + \text{Tr} [B] = \text{Tr} |\omega| = \|\omega\|_1. \quad (\text{B4})$$

In order to complete the proof, we need to show that $\bar{\Lambda}$ is the maximizing operator among all possible $-\mathbf{1} \leq \Lambda \leq \mathbf{1}$. First observe, by diagonalising A and B , that

$$\text{Tr} [\Lambda A] = \sum_k \alpha_k \langle a_k | \Lambda | a_k \rangle \leq \sum_k \alpha_k \langle a_k | a_k \rangle = \text{Tr} [A] \quad (\text{B5})$$

$$\text{Tr} [\Lambda B] = \sum_k \beta_k \langle b_k | \Lambda | b_k \rangle \geq \sum_k \beta_k (-\langle b_k | b_k \rangle) = -\text{Tr} [B]. \quad (\text{B6})$$

Thus

$$\text{Tr} [\Lambda \omega] = \text{Tr} [\Lambda A] - \text{Tr} [\Lambda B] \leq \text{Tr} [A] + \text{Tr} [B] = \text{Tr} |\omega| = \|\omega\|_1. \quad (\text{B7})$$

□

Proof of Lemma 1. Consider that

$$2D(\rho, \sigma) = \|\rho - \sigma\|_1 = \max_{-\mathbf{1} \leq \Lambda \leq \mathbf{1}} \text{Tr} [\Lambda(\sigma - \rho)] \quad (\text{B8})$$

$$\geq \text{Tr} [E(\sigma - \rho)], \quad (\text{B9})$$

which follows from applying Lemma 3 and from the fact that $0 \leq E \leq \mathbf{1}$ surely is one of the operators included in the maximization procedure. The result (27) is then easily obtained by separating the trace and rearranging terms in the previous inequality. \square

Proof of Lemma 2. Consider that

$$2D(\sqrt{E}\rho\sqrt{E}, \rho) = \|\rho - \sqrt{E}\rho\sqrt{E}\|_1 \leq \|\rho - \sqrt{E}\rho\|_1 + \|\sqrt{E}\rho - \sqrt{E}\rho\sqrt{E}\|_1 \quad (\text{B10})$$

$$= \left\| \left(\mathbf{1} - \sqrt{E} \right) \sqrt{\rho} \cdot \sqrt{\rho} \right\|_1 + \left\| \sqrt{E} \cdot \rho \left(\mathbf{1} - \sqrt{E} \right) \right\|_1. \quad (\text{B11})$$

thanks to the triangular inequality for the trace distance. Now for the first term write $\sqrt{\rho}$ in diagonal form $\{\sqrt{\lambda_k}, |f_k\rangle\}$ and use again the triangular inequality for the trace norm:

$$\left\| \left(\mathbf{1} - \sqrt{E} \right) \sqrt{\rho} \sum_k \sqrt{\lambda_k} |f_k\rangle \langle f_k| \right\|_1 \leq \sum_k \sqrt{\lambda_k} \left\| \left(\mathbf{1} - \sqrt{E} \right) \sqrt{\rho} |f_k\rangle \langle f_k| \right\|_1 \quad (\text{B12})$$

$$= \sum_k \sqrt{\lambda_k} \text{Tr} \sqrt{|f_k\rangle \langle f_k| \sqrt{\rho} \left(\mathbf{1} - \sqrt{E} \right)^2 \sqrt{\rho} |f_k\rangle \langle f_k|} \quad (\text{B13})$$

$$= \sum_k \sqrt{\lambda_k} \sqrt{\langle f_k | \sqrt{\rho} \left(\mathbf{1} - \sqrt{E} \right)^2 \sqrt{\rho} |f_k\rangle}. \quad (\text{B14})$$

Apply then the Cauchy-Schwarz inequality

$$|\vec{x} \cdot \vec{y}|^2 \leq |\vec{x}|^2 \cdot |\vec{y}|^2, \quad (\text{B15})$$

with $x_k = \sqrt{\lambda_k}$ and $y_k = \sqrt{\langle f_k | \sqrt{\rho} \left(\mathbf{1} - \sqrt{E} \right)^2 \sqrt{\rho} |f_k\rangle}$, to obtain

$$\left\| \left(\mathbf{1} - \sqrt{E} \right) \sqrt{\rho} \sum_k \sqrt{\lambda_k} |f_k\rangle \langle f_k| \right\|_1 \leq \sqrt{\sum_k \lambda_k \sum_j \langle f_j | \sqrt{\rho} \left(\mathbf{1} - \sqrt{E} \right)^2 \sqrt{\rho} |f_j\rangle} \quad (\text{B16})$$

$$\leq \sqrt{\text{Tr} \left[\rho \left(\mathbf{1} - \sqrt{E} \right)^2 \right]}, \quad (\text{B17})$$

where we used the fact that $\text{Tr}[\rho] = \sum_k \lambda_k \leq 1$. For the second term in (B11) write instead \sqrt{E} in its diagonal form $\{\sqrt{\nu_k}, |e_k\rangle\}$ and proceed in a similar way as before:

$$\left\| \sum_k \sqrt{\nu_k} |e_k\rangle \langle e_k| \rho (\mathbf{1} - \sqrt{E}) \right\|_1 \leq \sum_k \sqrt{\nu_k} \left\| |e_k\rangle \langle e_k| \rho (\mathbf{1} - \sqrt{E}) \right\|_1 \quad (\text{B18})$$

$$= \sum_k \sqrt{\nu_k} \left\| (\mathbf{1} - \sqrt{E}) \rho |e_k\rangle \langle e_k| \right\|_1 \quad (\text{B19})$$

$$= \sum_k \sqrt{\nu_k} \sqrt{\langle e_k| \rho (\mathbf{1} - \sqrt{E})^2 \rho |e_k\rangle} \quad (\text{B20})$$

$$\leq \sqrt{\sum_k \nu_k \sum_j \langle e_j| \rho (\mathbf{1} - \sqrt{E})^2 \rho |e_j\rangle} \quad (\text{B21})$$

$$\leq \sqrt{\text{Tr} \left[\rho^2 (\mathbf{1} - \sqrt{E})^2 \right]} \quad (\text{B22})$$

$$\leq \sqrt{\text{Tr} \left[\rho (\mathbf{1} - \sqrt{E})^2 \right]}, \quad (\text{B23})$$

where we used the triangular inequality, the invariance of the trace norm under hermitian conjugation, the Cauchy-Schwarz inequality, the fact that $\text{Tr}[E] = \sum_k \nu_k \leq 1$ and the property $\rho^2 \leq \rho \leq \mathbf{1}$. The inequality (B11) then simply becomes

$$2D(\sqrt{E}\rho\sqrt{E}, \rho) \leq 2\sqrt{\text{Tr} \left[\rho (\mathbf{1} - \sqrt{E})^2 \right]} \quad (\text{B24})$$

$$\leq 2\sqrt{\text{Tr} [\rho (\mathbf{1} - E)]}, \quad (\text{B25})$$

since

$$0 \leq E \leq \sqrt{E} \leq \mathbf{1} \quad (\text{B26})$$

$$\rightarrow (\mathbf{1} - \sqrt{E})^2 = \mathbf{1} + E - 2\sqrt{E} \leq \mathbf{1} - E. \quad (\text{B27})$$

Eventually we take the code average of (B25) and use the concavity of the square-root function and the hypothesis (28) to obtain the thesis (29):

$$\left\langle 2D(\sqrt{E}\rho\sqrt{E}, \rho) \right\rangle \leq 2\sqrt{\langle \text{Tr} [\rho (\mathbf{1} - E)] \rangle} \leq 2\sqrt{\epsilon}. \quad (\text{B28})$$

□

Appendix C: Derivation of the bisection POVM

Here we provide an explicit derivation of the POVM (40) associated with our bisection protocol. We consider each step to be carried out as a unitary process on an enlarged system, consisting of the state $|\Psi\rangle$ received by Bob (we take it pure for simplicity) and various ancillae, one for each step. The ancillae start in a reference state $|a\rangle$ and will turn into one of three possible states depending on the result of the measurement. In particular at the u -th step the ancilla state $|0\rangle$ ($|1\rangle$) corresponds to having found the codeword in set $\mathbf{C}_{k_1, \dots, k_{u-1}, 0}^{(u)}$ ($\mathbf{C}_{k_1, \dots, k_{u-1}, 1}^{(u)}$), while the state $|null\rangle$ corresponds to failure.

We start by applying the first-step POVM $\mathcal{M}^{(1)} = \{N_0^{(1)}, N_1^{(1)}, N_{null}^{(1)}\}$:

$$U^{(1)}(|\Psi\rangle|a\rangle_1) = \sqrt{N_0^{(1)}}|\Psi\rangle|0\rangle_1 + \sqrt{N_1^{(1)}}|\Psi\rangle|1\rangle_1 + \sqrt{N_{null}^{(1)}}|\Psi\rangle|null\rangle_1. \quad (C1)$$

After the second step POVM $\mathcal{M}^{(2)}$ we obtain the state

$$\begin{aligned} U^{(2)}(U^{(1)}(|\Psi\rangle|a\rangle_1)|a\rangle_2) &= \sqrt{N_{00}^{(2)}}\sqrt{N_0^{(1)}}|\Psi\rangle|0\rangle_1|0\rangle_2 + \sqrt{N_{01}^{(2)}}\sqrt{N_0^{(1)}}|\Psi\rangle|0\rangle_1|1\rangle_2 \\ &+ \sqrt{N_{10}^{(2)}}\sqrt{N_1^{(1)}}|\Psi\rangle|1\rangle_1|0\rangle_2 + \sqrt{N_{11}^{(2)}}\sqrt{N_1^{(1)}}|\Psi\rangle|1\rangle_1|1\rangle_2 \\ &+ \sqrt{N_{null}^{(2)}}\sqrt{N_0^{(1)}}|\Psi\rangle|0\rangle_1|null\rangle_2 + \sqrt{N_{null}^{(2)}}\sqrt{N_1^{(1)}}|\Psi\rangle|1\rangle_1|null\rangle_2 \\ &+ \sqrt{N_{null}^{(1)}}|\Psi\rangle|null\rangle_1|a\rangle_2. \end{aligned} \quad (C2)$$

If we stop at this step, the probability of having found $|\Psi\rangle$ in a given set, e.g. $\mathbf{C}_{10}^{(2)}$, is

$$P(10|\rho_\Psi) = \langle\Psi|\sqrt{N_1^{(1)}}N_{10}^{(2)}\sqrt{N_1^{(1)}}|\Psi\rangle = \text{Tr} \left[\left| \sqrt{N_{10}^{(2)}}\sqrt{N_1^{(1)}} \right|^2 \rho_\Psi \right], \quad (C3)$$

which corresponds to equation (40) for $\vec{k} = (1, 0)$ and can be easily generalized to an arbitrary number of steps.

REFERENCES

- ¹M. M. Wilde, *Quantum Information Theory* (Cambridge University Press 2013).
- ²A. S. Holevo, *Quantum Systems, Channels, Information* (de Gruyter Studies in Mathematical Physics, 2012).
- ³A. S. Holevo, Probl. Peredachi Inf. **9**, 3 (1973); Probl. Inf. Transm. (Engl. Transl.) **9**, 110 (1973).

- ⁴A. S. Holevo, IEEE Trans. Inf. Theory **44**, 269 (1998).
- ⁵B. Schumacher and M. D. Westmoreland, Phys. Rev. A **56**, 131 (1997); P. Hausladen, R. Jozsa, B. W. Schumacher, M. Westmoreland, and W. K. Wootters, *ibid.* **54**, 1869 (1996).
- ⁶P. Hausladen and W. K. Wootters, J. Mod. Opt. **41**, 2385 (1994).
- ⁷A. S. Holevo, e-print arXiv:quant-ph/9809023 [see also Tamagawa University Research Review, no. 4] (1998).
- ⁸A. Winter, IEEE Trans. Inf. Theory **45**, 2481 (1999).
- ⁹T. Ogawa, Ph.D. dissertation, University of Electro- Communications, Tokyo, Japan, 2000; (in Japanese) T. Ogawa and H. Nagaoka, in *Proceedings of the 2002 IEEE International Symposium on Information Theory*, Lausanne, Switzerland, (IEEE, New, York, 2002), p. 73; T. Ogawa, IEEE Trans. Inf. Theory **45**, 2486 (1999).
- ¹⁰T. Ogawa and H. Nagaoka, IEEE Trans. Inf. Theory **53**, 2261 (2007).
- ¹¹M. Hayashi and H. Nagaoka, IEEE Trans. Inf. Theory **49**, 1753 (2003).
- ¹²M. Hayashi, Phys. Rev. A **76**, 062301 (2007); Commun. Math. Phys. **289**, 1087 (2009).
- ¹³S. Lloyd, V. Giovannetti, L. Maccone, Phys. Rev. Lett. **106**, 250501 (2011).
- ¹⁴V. Giovannetti, S. Lloyd and L. Maccone, Phys. Lett. A **85**, 012302 (2012).
- ¹⁵P. Sen, e-print arXiv:1109.0802v1 [quant-ph] (2011).
- ¹⁶M. B. Hastings, Nat. Phys. **5**, 255 (2008).
- ¹⁷T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).
- ¹⁸B. Schumacher, Phys. Rev. A **51**, 2738 (1995).
- ¹⁹F. Hiai and D. Petz, Commun. Math. Phys. **143**, 99 (1991); T. Ogawa and H. Nagaoka, IEEE Trans. Inf. Theory **46**, 2428 (2000).
- ²⁰M. M. Wilde, S. Guha, *Proceedings of the 2012 International Symposium on Information Theory and its Applications*, 303-307 (2012).
- ²¹M. M. Wilde, S. Guha, S.-H. Tan, S. Lloyd, *Proceedings of the 2012 IEEE International Symposium on Information Theory* (ISIT 2012, Cambridge, MA, USA), 551-555.
- ²²A. S. Holevo and R. Werner, Phys. Rev. A **63**, 032312 (2001).
- ²³V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, J. H. Shapiro, and H. P. Yuen, Phys. Rev. Lett. **92**, 027902 (2004).
- ²⁴M. M. Wilde, Saikat Guha, IEEE Trans. Inf. Theory **59**, 1175 (2013).
- ²⁵E. Arikan, IEEE Trans. Inf. Theory **55**, 3051 (2009).

- ²⁶M. M. Wilde, O. Landon-Cardinal and P. Hayden, in *8th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2013)*, Dagstuhl, Germany, 2013 arXiv:1302.0398v1.
- ²⁷M. M. Wilde and J. M. Renes, in *Proceedings of the 2012 International Symposium on Information Theory and its Applications*, Honolulu, Hawaii, USA, October 2012. arXiv:1203.5794.
- ²⁸M. M. Wilde and J. M. Renes, in *Proceedings of the 2012 International Symposium on Information Theory*, Boston, Massachusetts, USA, July 2012. arXiv:1201.2906.
- ²⁹J. M. Renes and M. M. Wilde, IEEE Trans. Inf. Theory **60**, 3090 (2014).
- ³⁰M. M. Wilde, Proceedings of the Royal Society A **469**, 2157 (2013).
- ³¹The code average in the first two properties can be removed when using a stronger notion of conditional typicality, which we do not state here for simplicity, since such average will appear quite naturally during calculations when making use of Shannon’s averaging trick.
- ³²C. M. Caves and P. D. Drummond, Rev. Mod. Phys. **66**, 481 (1994).
- ³³S. L. Braunstein and P. van Loock, Rev. Mod. Phys. **77**, 513 (2005).
- ³⁴C. Weedbrook, S. Pirandola, R. Garcia-Patron, N. J. Cerf, T. C. Ralph, J. H. Shapiro and S. Lloyd, Rev. Mod. Phys. **84**, 621 (2012).
- ³⁵A. Winter, PhD Thesis, Universität Bielefeld (1999).
- ³⁶A. S. Holevo and V. Giovannetti, Rep. Prog. Phys. **75**, 046001 (2012).
- ³⁷R. G. Gallager, *Information Theory and Reliable Communication* (John Wiley & Sons 1968).
- ³⁸An alternative computation of the error probability can be carried out by employing a generalization of Sen’s non-commutative union bound³⁰, obtaining the same condition on the single-step POVM for the optimality of the bisection protocol.

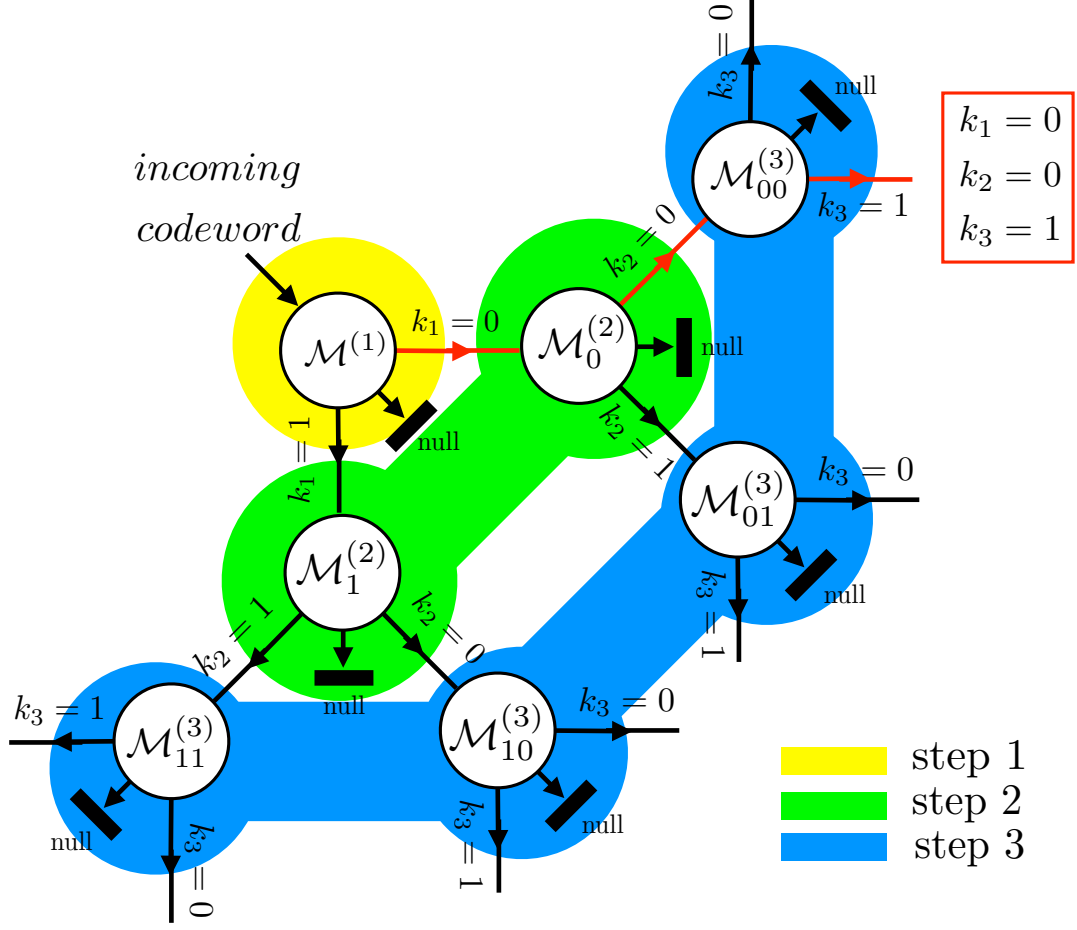


FIG. 1. Schematic representation of the bisection decoding procedure. It consists in a sequence of adaptive measurements which are performed in series of u_F concatenated steps, each being characterized by a POVM (the white circles) which admits three possible outcomes: two being associated respectively to the identification of the corresponding bit as 0 or 1, and one, the *null* outcome, associated with the event where no decision can be made on the value of the bit. The POVM to be performed at the u -th step depends upon the value of the bit obtained at the previous ones: for instance at the step number 2 Bob will perform either the POVM $\mathcal{M}_0^{(2)}$ or the POVM $\mathcal{M}_1^{(2)}$ depending on the value of k_1 he has obtained at the first step of the procedure, while at the step number 3 Bob will perform the POVMs $\mathcal{M}_{00}^{(3)}$, $\mathcal{M}_{01}^{(3)}$, $\mathcal{M}_{10}^{(3)}$, or $\mathcal{M}_{11}^{(3)}$ depending on the values of k_1 and k_2 obtained in the previous two steps. The figure refers to the case of $u_F = 3$, the redline representing the trajectory which yields Bob to assign the binary string $\vec{k} = (0, 0, 1)$ to the received codeword.

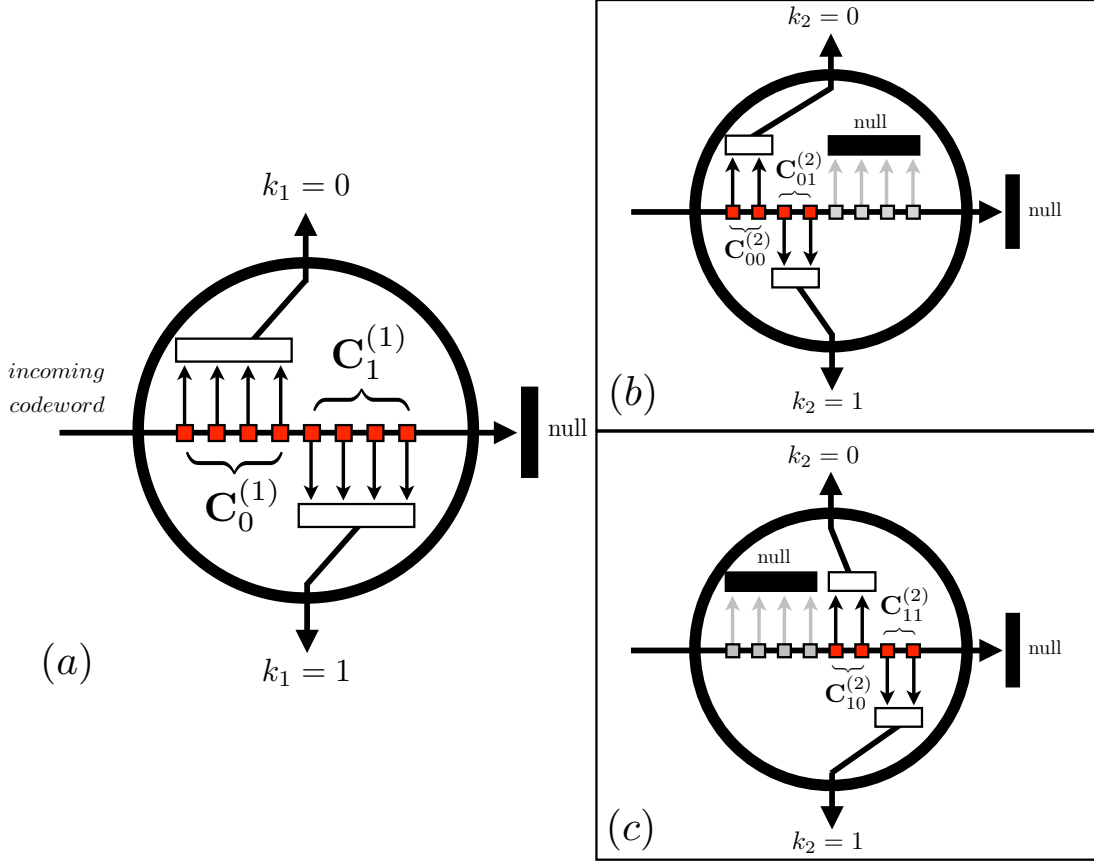


FIG. 2. Schematic representation of the set POVMs $\mathcal{M}^{(1)}$, $\mathcal{M}_0^{(2)}$, and $\mathcal{M}_1^{(2)}$ in terms of the sequential POVM decoding procedure (little square elements) for $N = 8$ codewords. Panel (a): implementation of $\mathcal{M}^{(1)}$. The red color of the square blocks indicates that all the elements of the sequential decoding POVM are active: their outcomes are used to determine whether the incoming codeword belongs to the subset $C_0^{(1)}$ (first four codewords), or to the subset $C_1^{(1)}$ (last four codewords) fixing the value of k_1 . The rectangular elements of the figure indicate that no other information is extracted from the outcomes of the sequential measurement. Panel (b): implementation of $\mathcal{M}_0^{(2)}$ which discriminates between the subsets $C_{00}^{(2)}$ and $C_{01}^{(2)}$. This element operates on the state emerging from the port $k_1 = 0$ of $\mathcal{M}^{(1)}$, see e.g. Fig. 1. As indicated by the color, only the first elements of the sequential POVM are active, while the outputs of the remaining ones are equivalent to the null result. Panel (c): implementation of $\mathcal{M}_1^{(2)}$ which discriminates among $C_{10}^{(2)}$ and $C_{11}^{(2)}$.